

FOR OFFICIAL USE ONLY

ACCESS DB # _____

PLEASE PRINT CLEARLY

Location (Bldg/Room#): _____

Scientific and Technical Information Center

SEARCH REQUEST FORM

Date: _____ Requester's Full Name: _____ Examiner #: _____

Art Unit: _____ Phone (301) _____ Serial Number: _____

Results Format Preferred (circle): PAPER DISK E-MAIL

To ensure an efficient and quality search, please attach a copy of the cover sheet, claims, and abstract or fill out the following:

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Date: _____

Search Topic:

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known.

For Sequence Searches Only Please include all pertinent information (parent, grandchild, divisional, or issued patent numbers) along with the appropriate serial number.

STAFF USE ONLY

Searcher: Beverly C 4994 _____ Type of Search _____ NA Sequence (#)

Searcher Phone #: _____ AA Sequence (#)

Searcher Location: _____ Structure (#)

Date Searcher Picked Up: _____ Bibliographic

Date Completed: 02-18-00 _____ Litigation

Searcher Prep & Review Time _____ Fulltext

Online Time: _____ Other

Vendors and Cost

_____ STN _____ Dialog

_____ Questel/Orbit _____ Dr. Link






_____ Lexis/Nexis _____ Westlaw

_____ WWW/Internet

_____ In-house sequence systems (list)

_____ Other (specify) MP

THIS PAGE BLANK (USPTO)

Release 3.1A John F. Collins, Biocomputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```
Run on:      Fri Feb 18 11:55:43 2000;  MasPar time 16.71 Seconds
            524.338 Million cell updates/sec
Abular output not generated.
```

```

Title:
>US-08-755-584-2
Description:
(1-412) from 5658711.pep
Perfect Score:
2858
Sequence:
1 MAEELRGGRGRSRKGGRRAR.....GHELPADGLSSGDLIEWG 412

```

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 08
Listing first 45 summaries

Database:

1:par1 2:par2 3:par3 4:par4 5:par5 6:par6 7:par7
8:par8 9:par9 10:par10 11:par11 12:par12 13:par13
14:par14 15:par15 16:par16 17:par17 18:par18
19:par19 20:par20 21:par21 22:par22 23:par23
24:par24 25:par25 26:par26 27:par27 28:par28
29:par29 30:par30 31:par31 32:par32 33:par33
34:par34 35:par35 36:par36 37:par37 38:par38
39:par39

Statistics: Mean 35.338; Variance 185.169; scale 0.191

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No
1	2858	100.0	412	28	M34091	Mouse NF-AT interact	2.68e-21
2	151	5.3	101	32	W60079	Homo sapiens sentrin-	3.49e-02
3	152	5.3	126	38	M87497	Human SMT3-like prote	2.99e-02
4	149	5.2	102	37	W74776	Human secreted protei	4.74e-02
5	122	4.3	265	35	W79128	FLGA annex stabilis1	2.74e+00
6	113	4.0	673	21	W09430	Human FXR2 polypheni	1.0e+00
7	109	3.8	1184	61	R78519	Cardiac annex1 cycl	1.77e+01
8	109	3.8	1184	6	R32882	Cardiac adenylyl cycl	1.77e+01
9	106	3.7	122	35	W68479	HIV-1 Strain YB30 re	2.70e+01
10	107	3.7	3248	18	R97975	Kinechoread protein C	2.35e+01
11	103	3.6	228	13	R72381	S. laevenduriae ORF3 ge	4.11e+01
12	102	3.6	788	25	W05393	Mouse SH3P2 protei.n	4.71e+01
13	102	3.6	2205	15	R79048	Infectious rubella vi	4.71e+01
14	99	3.5	213	24	W14567	Streptococcus pneumo	7.12e+01
15	100	3.5	386	22	W08664	Pneumogent human NF-H	6.21e+01
16	99	3.5	397	20	W09048	Plasmid pBBV DN AA en	7.12e+01

17	99	5	568	12	R66455	Serine-rich AF-pro	7.12e+01	
16	99	3	589	34	M60667	E.coli cold shock pro	7.12e+01	
18	99	3	5	1822	24	W26327	Human alpha-1 collagen	7.12e+01
20	100	3	5	1822	5	R21745	Extracellular factor	6.21e+01
21	98	3	30	23	W16376	Synthetic peptide 5.	8.16e+01	
22	97	3	123	13	R74990	Epstein-Barr virus nu	9.34e+01	
23	97	3	123	10	R51053	Epstein-Barr nuclear	9.34e+01	
24	97	3	166	7	R35450	Human eps8.	9.34e+01	
25	97	3	284	30	W34525	Ham sapiens tumour n	9.34e+01	
26	98	3	309	25	W23362	Mycobacterium tubercu	8.16e+01	
27	98	3	309	35	W64302	Mycobacterium tubercu	8.16e+01	
28	98	3	309	37	W81665	M. tuberculosis immu	8.16e+01	
29	96	3	359	31	W59829	Human I-kappa-B-beta	1.07e+02	
30	96	3	359	31	W56325	Amino acid sequence o	1.07e+02	
31	96	3	403	17	R85857	EBV nuclear antigen p	1.07e+02	
32	96	3	411	37	W76827	Human TR6 protein.	1.07e+02	
33	96	3	411	36	W79083	Human death domain co	1.07e+02	
34	98	3	450	36	W72145	HSV-2 strain DBS Cont	8.16e+01	
35	97	3	623	34	M68095	Chlamydomonas reinhar	9.34e+01	
36	96	3	927	2	P70768	Epstein-Barr virus gI	1.07e+02	
37	96	3	2466	35	W75999	Intracellular protein	1.07e+02	
38	96	3	2466	13	R71498	Human protein tyrosin	1.07e+02	
39	97	3	3433	23	W22017	Utrrophin.	9.34e+01	
40	95	3	396	30	W55819	Streptomyces roseofil	1.22e+02	
41	95	3	407	39	W95400	M. tuberculosis RP-fa	1.22e+02	
42	94	3	453	13	R73918	CD protein of Birmham	1.40e+02	
43	95	3	468	34	W64483	Human DR4 protein.	1.22e+02	
44	94	3	875	23	W19919	Human Ksr-1 (kinase s	1.40e+02	
45	94	3	1618	5	R21205	Human nestin.	1.40e+02	

ALIGNMENTS

ID	RESULT	1
AC	W34091 standard; Protein; 412 AA.	
AC	W34091;	
DT	18-MAY-1998 (first entry)	
DE	Mouse NF-AT interacting protein 45.	
KW	NF-AT interacting Protein 45; NIP45; yeast two-hybrid assay; mouse;	
KW	cell; transcription factor; cancer;	
KW	interleukin-4; IL-4; development; Th1; Th2; cytokine; allergy;	
KW	autoimmune disease; transplantation.	
OS	Mus sp.	
FH	Key	Location/Qualifiers
FT	Region	6..37
FN	W09739721-A2.	/note="highly basic region"
PD	30-OCT-1997.	
PE	23-APR-1997; 006708.	
PR	25-NOV-1996; US-755592.	
PR	23-APR-1996; US-636602.	
PR	25-NOV-1996; US-755584.	
PA	(HARD) HARVARD COLLEGE.	
PI	Glitcher LH, Ho I, Hodge MR;	
PI	WPI: 97-535556/49.	
DR	N-RSD5; T93045.	
PT	Production of cytokine(s) associated with Th2-type helper T cells -	
PT	particularly for controlling development of Th1 and Th2 cells for	
PT	treatment of allergy, autoimmune disease etc.	
PS	Claim 31: Fig 11: 151pp; English.	
CC	This is the amino acid sequence of the mouse NF-AT interacting Protein 45	
CC	(NIP45). The gene sequence was isolated by using a yeast two-hybrid	
CC	detection system for proteins that interact with the NF-AT Rel homology	
CC	domain (RHD). The assay used, as a "bait", a 900 bp fragment of the	
CC	mouse NF-ATp encoding the region spanning amino acids 228-250. NF-AT is	
CC	a multisubunit transcription complex containing a cyclosporin A sensitive	
CC	cytoplasmic phosphoprotein and an inducible component of the Ap-1 family	
CC	of transcription factors. The screen was carried out on a cDNA library	
CC	prepared from the murine T cell line D10. One class of proteins, NIP45	
CC	designated NIP45, bound the NF-AT-RHD region with high affinity.	
CC	CC can be used in a claimed method to inhibit or stimulate production of	
CC	NF-AT family protein, particularly interleukin-4 (IL-4) and the treated	
CC	cells may be administered to control development of Th1 or Th2 cells by	

Query Match	Similarity	Score	DB	Length
Best Local	31.6%	Pred. No. 2,99e-02;		
Matches	25;	Conservative	19;	Mismatches 33; Indels 2; Gaps 2

ID	Accession	Protein	Length
Dd	8	evgkxendh1nk1kvaggdgs-vvqfkl1khrplek1mkaycergqlsmrgtfrfdgqpl 66	
Oy	333	ENTERSQE-LKLRVQGRKKHOMLEISLSPDPLKVLVMSHYEAMGLSGHKLSPFFDGTKL 391	
Dd	67	netdtpaqglemededtlidv 85	
Oy	392	SGKELPADLGEISDLEIV 410	

RESULT	Accession	Protein	Length
ID	W74776	standard; Protein; 102 AA.	
AC	W74776;		
DT	25-JAN-1999	(first entry)	
DE	Human secreted protein encoded by gene 47, clone HOGAN75.		
HN	Human; secreted protein; testis; tumour; foetal brain tissue;		
KW	fusion protein; cancer; central nervous system; seizure;		
DI	diagnosis; neurodegenerative disease.		
OS	Homo sapiens.		
FT	Key	Location/Qualifiers	
FT	Misc-difference	102	
FT		label= unknown	
PD	WO9839448-A2.		
PD	11-SEP-1998.		
PE	06-MAR-1998;	U04493.	
PE	02-OCT-1997;	US-061060.	
PR	07-MAR-1997;	US-038621.	
PR	07-MAR-1997;	US-040161.	
PR	07-MAR-1997;	US-040162.	
PR	07-MAR-1997;	US-040163.	
PR	07-MAR-1997;	US-040333.	
PR	07-MAR-1997;	US-040334.	
PR	07-MAR-1997;	US-040336.	
PR	07-MAR-1997;	US-040626.	
PR	11-APR-1997;	US-043311.	
PR	11-APR-1997;	US-043312.	
PR	11-APR-1997;	US-043313.	
PR	11-APR-1997;	US-043314.	
PR	11-APR-1997;	US-043568.	
PR	11-APR-1997;	US-043569.	
PR	11-APR-1997;	US-043576.	
PR	11-APR-1997;	US-043578.	
PR	11-APR-1997;	US-043580.	
PR	11-APR-1997;	US-043669.	
PR	11-APR-1997;	US-043670.	
PR	11-APR-1997;	US-043671.	
PR	11-APR-1997;	US-043672.	
PR	11-APR-1997;	US-043674.	
PR	23-MAY-1997;	US-047492.	
PR	23-MAY-1997;	US-047500.	
PR	23-MAY-1997;	US-047501.	
PR	23-MAY-1997;	US-047502.	
PR	23-MAY-1997;	US-047503.	
PR	23-MAY-1997;	US-047581.	
PR	23-MAY-1997;	US-047582.	
PR	23-MAY-1997;	US-047583.	
PR	23-MAY-1997;	US-047584.	
PR	23-MAY-1997;	US-047585.	
PR	23-MAY-1997;	US-047586.	
PR	23-MAY-1997;	US-047587.	
PR	23-MAY-1997;	US-047588.	
PR	23-MAY-1997;	US-047589.	
PR	23-MAY-1997;	US-047590.	
PR	23-MAY-1997;	US-047592.	
PR	23-MAY-1997;	US-047593.	
PR	23-MAY-1997;	US-047594.	
PR	23-MAY-1997;	US-047595.	
PR	23-MAY-1997;	US-047596.	
PR	23-MAY-1997;	US-047597.	
PR	23-MAY-1997;	US-047598.	
PR	23-MAY-1997;	US-047599.	

PR	23-MAY-1997	US-047599.
PR	23-MAY-1997	US-047600.
PR	23-MAY-1997	US-047601.
PR	23-MAY-1997	US-047612.
PR	23-MAY-1997	US-047613.
PR	23-MAY-1997	US-047614.
PR	23-MAY-1997	US-047615.
PR	23-MAY-1997	US-047617.
PR	23-MAY-1997	US-047618.
PR	23-MAY-1997	US-047632.
PR	23-MAY-1997	US-047633.
PR	06-JUN-1997	US-048964.
PR	06-JUN-1997	US-048974.
PR	13-JUN-1997	US-049610.
PR	08-JUL-1997	US-051926.
PR	16-JUL-1997	US-052874.
PR	18-AUG-1997	US-055724.
PR	22-AUG-1997	US-056630.
PR	22-AUG-1997	US-056631.
PR	22-AUG-1997	US-056632.
PR	22-AUG-1997	US-056636.
PR	22-AUG-1997	US-056637.
PR	22-AUG-1997	US-056662.
PR	22-AUG-1997	US-056664.
PR	22-AUG-1997	US-056845.
PR	22-AUG-1997	US-056862.
PR	22-AUG-1997	US-056864.
PR	22-AUG-1997	US-056872.
PR	22-AUG-1997	US-056874.
PR	22-AUG-1997	US-056879.
PR	22-AUG-1997	US-056880.
PR	22-AUG-1997	US-056881.
PR	22-AUG-1997	US-056882.
PR	22-AUG-1997	US-056884.
PR	22-AUG-1997	US-056886.
PR	22-AUG-1997	US-056887.
PR	22-AUG-1997	US-056888.
PR	22-AUG-1997	US-056889.
PR	22-AUG-1997	US-056892.
PR	22-AUG-1997	US-056893.
PR	22-AUG-1997	US-056894.
PR	22-AUG-1997	US-056903.
PR	22-AUG-1997	US-056908.
PR	22-AUG-1997	US-056909.
PR	22-AUG-1997	US-056910.
PR	22-AUG-1997	US-056911.
PR	05-SEP-1997	US-057650.
PR	05-SEP-1997	US-057659.
PR	05-SEP-1997	US-057761.
PR	12-SEP-1997	US-058785.
PA	(HUMA-) HUMAN GENOME SCI INC.	
PI	Bednarik DP, Brewer LA, Carter KC, Duan R, Edner R, Endress GA,	
PI	Feng P, Fertle AM, Fischer CL, Florence KA, Greene JM, Hu JS,	
PI	Kyng H, Laflaur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,	
PI	Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;	
PI	WPI: 98-506364/43.	
DR	N-PSDB: V59557.	
PT	New isolated human genes and the secreted polypeptide(s) they encode	
PT	- useful for diagnosis and treatment of e.g. cancers, neurological	
PT	disorders, immune diseases, inflammation or blood disorders	
PS	Claim 1; Page 566-567; 721pp; English.	
CC	This sequence represents a secreted human protein encoded by the nucleic	
CC	acid molecule designated Gene 47 from the human cDNA clone HOGAV75	
CC	(deposited as clone ATCC 97899 and ATCC 209045).	
CC	The gene can be used to generate fusion proteins by linking to the gene	
CC	to a human immunoglobulin Fc portion (e.g. V59502) for increasing the	
CC	stability of the fused protein as compared to the human protein only.	
CC	The invention relates to 186 novel genes and their fragments (nucleic	
CC	acid sequences: V59511-V59812; amino acid sequences W74731-W75026) which	


```

CC      cyclase      1184 AA:
SQ      Sequence      1184 AA:

Query Match      3.8%; Score 109; DB 14; Length 1184;
Best Local Similarity 50.0%; Pred. No. 1.77e+01;
Matches      19; Conservative      9; Mismatches      8; Indels      2; Gaps      2

      84 apylggagpgraaagppprtaratrgrrirpryagat 121
      ||| ||| : ||| ||| : ||| ||| : ||| ||| : |||
      2 AEPLRGRRGPRSRGGRRARRARG-ARGRCPRAR-OSPAR 37

RESULT      8
ID      R32882 standard; Protein; 1184 AA.
AC      R32882.
DT      17-JUN-1993 (first entry)
DE      Cardiac adenylyl cyclase type V.
KW      CACV; therapy; diagnostic; cardiac function; cyclic AMP; cAMP; heart
KW      failure.
OS      Canis familiaris.
EP      E529622-A.
EP      03-MAR-1993.
PR      27-AUG-1992; 114637.
PR      29-AUG-1991; US-751460.
PA      (AMCY ) AMERICAN CYANAMID CO.
PI      Ishikawa Y, Kanski AF;
PI      WPI: 93-068688/09.
DR      N-PSDB; 037543.
PT      Isolated nucleic acid mol. encoding Cardiac adenylyl cyclase type
PT      V - useful for determining and modifying cardiac function
PS      Claim 4; Page 15-27; 38pp; English.
CC      Left ventricular tissue of canine heart was used as a source of mRNA.
CC      A cDNA library was prepd. In lambda g10 phage. A 970 bp Aat-HincII
CC      fragment from type I adenylyl cyclase cDNA was used as probe. The
CC      clones isolated were used to obtain cDNA encoding CACV. This probe
CC      may also be used to screen a human cardiac cDNA library to obtain
CC      the cDNA encoding human CACV. CACV, its analogues and antibodies
CC      are useful in therapy or diagnostic assays, e.g. in modifying and
CC      determining cardiac function. A decrease in CACV content of the
CC      heart contributes to impaired cAMP prodn. and in heart failure. The
CC      CACV can also be used to screen for cpds. which stimulate or inhibit
CC      the activity of the cyclase.
CC      Sequence      1184 AA:

Query Match      3.8%; Score 109; DB 6; Length 1184;
Best Local Similarity 50.0%; Pred. No. 1.77e+01;
Matches      19; Conservative      9; Mismatches      8; Indels      2; Gaps      2

      84 apylggagpgraaagppprtaratrgrrirpryagat 121
      ||| ||| : ||| ||| : ||| ||| : ||| ||| : |||
      2 AEPLRGRRGPRSRGGRRARRARG-ARGRCPRAR-OSPAR 37

RESULT      9
ID      W68479 standard; Protein; 122 AA.
AC      W68479.
DT      08-DEC-1998 (first entry)
DE      HIV-1 strain YBF30 rev protein.
KW      HIV-1 strain YBF30; antibody; oligonucleotide; diagnosis; immunisation;
KW      infection; typing; rev.
OS      Human immunodeficiency virus type 1.
FT      Key      Location/Qualifiers
FT      MISC_difference 104
FT      /label= unknown
FT      /note= "encoded by TGA"
PD      12-JUN-1998.
PF      09-DEC-1996; 015087.
PF      09-DEC-1996; FR-015087.
PA      (ASSI-) ASSISTANCE PUBLIQUE HOPITAUX PARIS.
PA      (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PI      Baire-Sinoussi F, Lousset-Ajaka I, Maucelere P, Saragosti S,

```

[illegible]

THIS PAGE BLANK (USPTO)

(\mathcal{M})

Distribution rights by Oxford Molecular Ltd

```
Run on: Fri Feb 18 12:02:02 2000; MasPar time 7.55 Seconds
```

[illegible]

1
2
3
4
5

Gap 11

31253 seqs, 12956647 residues

Minimum Match 08

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1

```
mean 32.940;  Variance 177.353;  scale 0.186
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	2858	100.0	412	2	US-08-75-	Sequence 2, Applicatio	2.34e+21	
2	160	5.6	104	2	US-08-83-	Sequence 3, Applicatio	4.58e-03	
3	152	5.3	126	2	US-08-83-	Sequence 1, Applicatio	1.57e-02	
4	113	4.0	673	2	US-08-45-	Sequence 6, Applicatio	5.18e+00	
5	106	3.7	265	2	US-08-97-	Sequence 1, Applicatio	1.39e+01	
6	107	3.7	3248	3	PCT-US93-1	Sequence 1, Applicatio	1.12e+01	
7	107	3.7	3248	1	US-08-33-	Sequence 1, Applicatio	1.12e+01	
8	102	3.6	2205	1	US-08-03-	Sequence 2, Applicatio	2.44e+01	
9	100	3.5	181	2	US-08-76-	Sequence 56, Applicatio	3.21e+01	
10	99	3.5	397	3	PCT-US94-0	Sequence 6, Applicatio	3.69e+01	
11	99	3.5	568	3	PCT-US94-0	Sequence 30, Applicatio	3.69e+01	
12	99	3.5	568	1	US-08-33-	Sequence 30, Applicatio	3.69e+01	
13	98	3.4	30	1	US-08-28-	Sequence 10, Applicatio	4.23e+01	
14	97	3.4	123	2	US-08-36-	Sequence 5, Applicatio	4.85e+01	
15	97	3.4	123	2	US-08-24-	Sequence 1, Applicatio	4.85e+01	
16	97	3.4	166	1	US-07-935-	Sequence 1, Applicatio	4.85e+01	
17	97	3.4	166	3	PCT-US93-0	Sequence 2, Applicatio	4.85e+01	
18	97	3.4	166	1	US-08-368-	Sequence 2, Applicatio	4.85e+01	
19	98	3.4	237	2	US-08-97-	Sequence 5, Applicatio	4.23e+01	
20	97	3.4	359	2	US-09-02-	Sequence 6, Applicatio	4.85e+01	
21	96	3.4	359	2	US-08-82-	Sequence 3, Applicatio	4.55e+01	
22	97	3.4	389	2	US-08-82-	Sequence 14, Applicatio	5.55e+01	
23	96	3.4	404	2	US-09-032-	Sequence 3, Applicatio	4.85e+01	

24	96	3.4	857	1	US-08-413-	Sequence 10,	Applicati	5.55e+01
25	96	3.4	857	1	US-08-220-	Sequence 10,	Applicati	5.55e+01
26	96	3.4	2465	2	US-08-596-	Sequence 3,	Applicati	5.55e+01
27	96	3.4	2465	2	PCT-US94-0	Sequence 2,	Applicati	5.55e+01
28	96	3.3	384	2	US-08-673-	Sequence 11,	Applicati	6.36e+01
29	95	3.3	384	2	US-08-614-	Sequence 11,	Applicati	6.36e+01
30	95	3.3	478	2	US-09-040-	Sequence 8,	Applicati	6.36e+01
31	94	3.3	635	1	US-08-571-	Sequence 10,	Applicati	7.28e+01
32	94	3.3	635	1	US-08-909-	Sequence 10,	Applicati	7.28e+01
33	94	3.3	635	1	US-08-909-	Sequence 10,	Applicati	7.28e+01
34	93	3.3	873	1	US-08-909-	Sequence 6,	Applicati	8.33e+01
35	93	3.3	873	1	US-08-909-	Sequence 6,	Applicati	8.33e+01
36	93	3.3	873	1	US-08-571-	Sequence 6,	Applicati	8.33e+01
37	94	3.3	875	1	US-08-571-	Sequence 8,	Applicati	7.28e+01
38	94	3.3	875	1	US-08-909-	Sequence 8,	Applicati	7.28e+01
39	94	3.3	875	1	US-08-909-	Sequence 8,	Applicati	7.28e+01
40	94	3.3	1112	2	US-08-714-	Sequence 2,	Applicati	7.28e+01
41	94	3.2	1618	1	US-07-853-	Sequence 4,	Applicati	9.55e+01
42	94	3.2	222	2	US-08-933-	Sequence 41,	Applicati	9.55e+01
43	92	3.2	351	1	US-08-646-	Sequence 2,	Applicati	9.55e+01
44	92	3.2	704	1	US-08-646-	Sequence 3,	Applicati	9.55e+01
45	92	3.2	1612	1	PCT-US94-0	Sequence 48,	Applicati	9.55e+01

ALIGNMENTS

RESULT	1	STANDARD;	PRT;	412 AA.
ID	US-08-755-584-2			

AC XXXXXXX

DT

DE Sequen

CC Sequences

CC GENE:

CC
APCC
TITCC
CO

88

88

88

22

22

CC
CU

188

CC PR:

22

22

33

22

CC
INFO

22

RESULT	1	STANDARD;	PRT;	412 AA.
ID	US-08-755-584-2			
XX	xxxxxx			
XX				
XX				
DE	Sequence 2, Application US/08755584			
CC	Sequence 2, Application US/08755584			
CC	Patent No. 5858711			
CC	GENERAL INFORMATION:			
CC	APPLICANT: Glimcher, Laurie H.			
CC	APPLICANT: Hodge, Martin R.			
CC	TITLE OF INVENTION: NF-AT-INTERACTING PROTEIN NIP45 AND METHODS			
CC	TITLE OF INVENTION: OF USE THEREFOR			
CC	NUMBER OF SEQUENCES: 2			
CC	CORRESPONDENCE ADDRESS:			
CC	ADDRESSEE: LAHIVE & COCKFIELD			
CC	STREET: 60 State Street, suite 510			
CC	CITY: Boston			
CC	STATE: Massachusetts			
CC	COUNTRY: USA			
CC	ZIP: 02109-1875			
CC	COMPUTER READABLE FORM:			
CC	MEDIUM TYPE: Floppy disk			
CC	COMPUTER: IBM PC compatible			
CC	OPERATING SYSTEM: PC-DOS/MS-DOS			
CC	SOFTWARE: PatentIn Release #1.0, Version #1.25			
CC	CURRENT APPLICATION DATA:			
CC	APPLICATION NUMBER: US/08/755,584			
CC	FILING DATE:			
CC	CLASSIFICATION: 435			
CC	PRIOR APPLICATION DATA:			
CC	APPLICATION NUMBER:			
CC	FILING DATE:			
CC	ATTORNEY/AGENT INFORMATION:			
CC	NAME: Giulio A. DeConti, Jr.			
CC	REGISTRATION NUMBER: 31,503			
CC	REFERENCE/DOCKET NUMBER: HUI-026			
CC	TELECOMMUNICATION INFORMATION:			
CC	TELEPHONE: (617)227-7400			
CC	TELEFAX: (617)227-5941			
CC	INFORMATION FOR SEQ ID NO: 2:			
CC	SEQUENCE CHARACTERISTICS:			
CC	LENGTH: 412 amino acids			
CC	TYPE: amino acid			

Fri Feb 18 13:12:00 2000

US-08-755-584-2.rai

Page 3

```
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC IMMEDIATE SOURCE:
CC LIBRARY: LUNGET03
CC CLONE: 2361410
CC SEQUENCE 126 AA; 13961 MW; 87522 CN;

Query Match          5.3%; Score 152; DB 2; Length 126;
Best Local Similarity 31.6%; Pred. No. 1,57e-02;
Matches      25; Conservative   19; Mismatches 33; Indels    2; Gaps    2;

Db       8 EGVKRNHINKVAGDGS-VVOFKIRKHPLSLTKMAYCROGLSMRQIFRRDGAPI 66
OY      333 EXTETSOE-LRLRVGKEKHOMLEISLSPDSPLKVMISHFEAMGLSGHKLSFFPDGTL 391
        | :|::|||:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       67 NETDTPAOLEMEDDETDIV 85
OY      : ||:| | | |:|
        392 SGKELPADLGESDLIEV 410

RESULT              4
US -08-455-073A-6 STANDARD: PRT; 673 AA.

XX      xxxxxx

Sequence 6, Application US/08455073A
Patent No. 5876949
GENERAL INFORMATION:
APPLICANT: Gideon Dreyfuss
APPLICANT: Mikiko C. Sioml
APPLICANT: Yan Zhang
TITLE OF INVENTION: Fragile X Related Proteins, Compositions And Methods
NUMBER OF INVENTIONS: Of Making And Using The Same
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & NO. 5876949rls
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/455,073A
FILING DATE: 31-MAY-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: DeLuca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: UPN-2201
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3430
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 673 amino acid residues
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
SEQUENCE 673 AA; 74079 MW; 2221704 CN;

Query Match          4.0%; Score 113; DB 2; Length 673;
Best Local Similarity 32.4%; Pred. No. 5.18e+00;
Matches     23; Conservative   18; Mismatches 26; Indels    4; Gaps    4;
```

```

Db 471 RDPTRGCESSRRRTGCRGPPAPPTSSSSISVLDKDDSPNPSLDTSSE-EP 529
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 8 RGRPSRGRGRARRRGARGR-CPNAROSPARLIDIVLVLY-SDSE-EYLEVADEVY 64

Db 530 PVDSEGEPPP 540
      || : : |
OY 65 PVARLPAPAKP 75

RESULT 5 STANDARD; PRT; 265 AA.
ID US-08-970-133-1
AC xxxxxx
AD
DT
PI
XX
XX
XX
Sequence 1, Application US/08970133
CC
CC Sequence 1, Application US/08970133
CC Patent No. 5916753
CC
CC GENERAL INFORMATION:
CC APPLICANT: Bandman, Olga
CC APPLICANT: Guegler, Karl J.
CC APPLICANT: Lal, Preeti
CC TITLE OF INVENTION: SH3-CONTAINING PROTEINS
CC NUMBER OF SEQUENCES: 5
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Incyte Pharmaceuticals, Inc.
CC STREET: 5174 Porter Dr.
CC CITY: Palo Alto
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94304
CC
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FastSeq for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/970,133
CC FILING DATE: Filed Herewith
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Billings, Lucy J.
CC REGISTRATION NUMBER: 36,749
CC REFERENCE/DOCKET NUMBER: PF-0419 US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 650-855-0555
CC TELEFAX: 650-845-4166
CC INFORMATION FOR SEQ. ID NO: 1:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 265 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC IMMEDIATE SOURCE:
CC LIBRARY: BRAITUT03
CC CLONE: 865744
SQ SEQUENCE 265 AA; 30106 MW; 347009 CN;

Query Match 3.7%; Score 106; DB 2; Length 265;
Best Local Similarity 28.6%; Pred. No. 1.3se+01;
Matches 22; Conservative 21; Mismatches 30; Indels 4; Gaps 4.

Db 38 NKGATPEFDSNLPE-ORRKKLQGVDELKKEIQKENDQDAITKMDVYLKNPOMGDPA 96
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 150 NSGSSPDEADALPGSGWRKKLRKKCKEKKKME-EFPDD-ISPDPSSRNKSKRHT 207
      . : | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
      97 SLDRKLAEVSONIEKR 113
      . : | | | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

QY 208 AL-OKLEVNKRLODLR 223

RESULT 6
ID PCT-US95-16216-1 STANDARD; PRT; 3248 AA.

XX AC xxxxxx

DE Sequence 1, Application PC/TUS9516216

XX Sequence 1, Application PC/TUS9516216

CC GENERAL INFORMATION:

CC APPLICANT: Yen, Timothy J.

CC APPLICANT: Rattner, Jerome B.

CC TITLE OF INVENTION: Nucleic Acid Encoding a Transiently

CC TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use

CC NUMBER OF SEQUENCES: 4

CC CORRESPONDENCE ADDRESSES:

CC ADDRESSEE: Dann, Dorfman, Herrell and Skillman

CC STREET: 1601 Market Street Suite 720

CC CITY: Philadelphia

CC STATE: PA

CC COUNTRY: USA

CC ZIP: 19103-2307

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patentin Release #1.0, Version #1.30

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: PCT/US95/16216

CC FILING DATE:

CC CLASSIFICATION:

CC PRIORITY APPLICATION DATA:

CC APPLICATION NUMBER: US 08/353,700

CC FILING DATE: 09-DEC-1995

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Reed, Janet E.

CC REGISTRATION NUMBER: 36,252

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (215) 563-4100

CC TELEFAX: (215) 563-4044

CC INFORMATION FOR SEQ ID NO: 1:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 3248 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: not relevant

CC TOPOLOGY: not relevant

CC MOLECULE TYPE: protein

CC HYPOTHETICAL: NO

CC ANTI-SENSE: NO

CC ORGANISM: HUMAN

SQ SEQUENCE 3248 AA; 372207 MW; 51689535 CN;

DB 31 KKEKQROFOLDSEAPQKOTQVENEKTEG-TNLKRENORLMEICESLEKTKOKISHE 89

QY 177 KEEKMEEFPPDQDISPLPQSSSRKRSRHTALOKLEVRNKRLODLNCSLS-PROHOSPA 235

DB 90 LQVRESQVNFQEGOLNSGKROIEKLEQELKRCRSELER 127

QY 236 LQSTDDEVVLVEGPLYLPOSSRLFTL-K-I-RCRADLYR 270

RESULT 7
ID US-08-353-700-1 STANDARD; PRT; 3248 AA.

XX AC xxxxxx

DT Sequence 1, Application US/08353700

XX Sequence 1, Application US/08353700

XX Patent No. 5599919

CC GENERAL INFORMATION:

CC APPLICANT: YEN, TIMOTHY J.

CC APPLICANT: RATTNER, JEROME B.

CC TITLE OF INVENTION: NUCLEIC ACID ENCODING A

CC TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,

CC TITLE OF INVENTION: AND METHODS OF USE

CC NUMBER OF SEQUENCES: 4

CC CORRESPONDENCE ADDRESSES:

CC ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN

CC STREET: 1601 MARKET STREET, SUITE 720

CC CITY: PHILADELPHIA

CC STATE: PA

CC COUNTRY: USA

CC ZIP: 19103-2307

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patentin Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/353,700

CC FILING DATE: 09-DEC-1994

CC CLASSIFICATION: 435

CC ATTORNEY/AGENT INFORMATION:

CC NAME: REED, JANET E.

CC REGISTRATION NUMBER: 36,252

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (215) 563-4100

CC TELEFAX: (215) 563-4044

CC INFORMATION FOR SEQ ID NO: 1:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 3248 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC HYPOTHETICAL: NO

CC ANTI-SENSE: NO

CC ORIGINAL SOURCE:

CC ORGANISM: HUMAN

SQ SEQUENCE 3248 AA; 372207 MW; 51689535 CN;

DB 31 KKEKQROFOLDSEAPQKOTQVENEKTEG-TNLKRENORLMEICESLEKTKOKISHE 89

QY 177 KEEKMEEFPPDQDISPLPQSSSRKRSRHTALOKLEVRNKRLODLNCSLS-PROHOSPA 235

DB 90 LQVRESQVNFQEGOLNSGKROIEKLEQELKRCRSELER 127

QY 236 LQSTDDEVVLVEGPLYLPOSSRLFTL-K-I-RCRADLYR 270

RESULT 8
ID US-08-093-453B-2 STANDARD; PRT; 2205 AA.

XX AC xxxxxx

DE Sequence 2, Application US/08093453B

CC Sequence 2, Application US/08093453B

CC Patent No. 5439816

CC GENERAL INFORMATION:

```
CC APPLICANT: Frey, Teryl K.  
CC APPLICANT: Dominguez, Geraldina  
CC APPLICANT: Wang, Chin_Yen  
CC TITLE OF INVENTION: Modified Infectious Rubella Virus  
CC NUMBER OF SEQUENCES: 13  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Jamie L. Greene, Jones & Askew  
CC STREET: 191 Peachtree Street, 37th Floor  
CC CITY: Atlanta  
CC STATE: Georgia  
CC COUNTRY: United States  
CC ZIP: 30303  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Diskette, 3.50  
CC COMPUTER: Macintosh  
CC OPERATING SYSTEM: 7.0  
CC SOFTWARE: Microsoft Word  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/093,453B  
CC FILING DATE: 19 JUL 1993  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: U.S. 07/722,334  
CC FILING DATE: 28 JUN 1991  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Greene, Jamie L.  
CC REGISTRATION NUMBER: 32,467  
CC REFERENCE/DOCKET NUMBER: 07362-0101  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 404_818-3700  
CC TELEFAX: 404_818-3799  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 2205 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: peptide  
CC HYPOTHEICAL: NO  
CC ANTI-SENSE: NO  
CC FRAGMENT TYPE: N_terminal  
CC ORIGINAL SOURCE:  
CC ORGANISM: Rubella virus  
CC STRAIN: Thierlen  
CC CC  
SQ SEQUENCE 2205 AA; 240235 MW; 22500019 CN;  
  
Query March 3.6%; Score 102; DB 1; Length 2205;  
Best Local Similarity 33.3%; Pred.No.2,44e+01;  
Matches 37; Conservative 17; Mismatches 48; Indels 9; Gaps .7  
  
D 2086 GARPPPGHHRRARQSDSP-LRAROSPFRRLTPLYGFLIPTLRSSPTVVSPLVTGTQ 2144  
I::| | | | | | | | | | | | | | | | | | | | | |  
Y 9 GPRSGGGGARAKARAGRCRPARQSPAR-LIP--DT-VLVLDVSDSEV-LEVADPV 62  
I::| | | | | | | | | | | | | | | | | | | | | |  
Db 2145 LLPGFRAIGCPNGFYYPHNG-GPPEGGRGTITPPARPATRRRLABAPAA 2194  
I::| | | | | | | | | | | | | | | | | | | | | |  
Y 63 EVPAARLPAPAKPEODSDSDEGAEGAGAG-PRTLVRRRRRLDDGEAP 112  
I::| | | | | | | | | | | | | | | | | | | | | |  
  
RESULT 9  
ID US-08-726-306A-56 STANDARD; PRT; 181 AA.  
XX xxxxxx  
DT  
XX  
XX Sequence 56, Application US/08726306A  
DE  
XX Sequence 56, Application US/08726306A  
CC Patent No. 5958684  
CC GENERAL INFORMATION:  
CC APPLICANT: van Leeuwen, Frederik Willem  
CC APPLICANT: Burbach, JohannesPeter Henri
```

```

CC      APPLICANT: Grosveld, Franklin G.
CC      TITLE OF INVENTION: DIAGNOSIS METHOD AND REAGENTS
CC      NUMBER OF SEQUENCES: 189
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE: Banner & Wilcoff, Ltd.
CC      STREET: 1 Financial Center
CC      CITY: Boston
CC      STATE: MA
CC      COUNTRY: US
CC      ZIP: 02111
CC      COMPUTER READABLE FORM:
CC      MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
CC      COMPUTER: IBM PC compatible
CC      OPERATING SYSTEM: PC-DOS/MS-DOS
CC      SOFTWARE: Wordperfect 6.1
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER: US/08/726,306A
CC      FILING DATE: 02-Oct-1996
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER: GB 95/20080.4
CC      FILING DATE: 02-Oct-1995
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER: US 60/009,832
CC      FILING DATE: 01-Jan-1996
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: Williams, Ph.D., Kathleen M.
CC      REGISTRATION NUMBER: 34,380
CC      REFERENCE/DOCKET NUMBER: 96,048-A (3255/00784)
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: (617) 345-9100
CC      TELEFAX: (617) 345-9111
CC      INFORMATION FOR SEQ ID NO: 56:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 181 amino acids
CC      TYPE: amino acid
CC      STRANDEDNESS: Single
CC      TOPOLOGY: unknown
CC      MOLECULE TYPE: peptide
CC      SEQUENCE 181 AA; 17613 MW; 98529 CN;
SQ
DB      Query Match          3.5%; Score 100; DB 2; Length 181;
DB      Best Local Similarity 53.2%; Pred. No. 3,21e+01;
DB      Matches 16; Conservative 6; Mismatches 5; Indels 2; Gaps 2.
QY      42 PLRYGRAVBARGRDARG- AAPRGARS 69
QY      ||| ||:|:| | |||: :| ||:|:
QY      4 PLR-GRGPRSRGGRARARARGARGRCPRA 31

RESULT 10
ID      PCT-US96-10602-6      STANDARD:      PRT: 397 AA.
XX      xxxxxx
XX      DT
XX      DE
XX      Sequence 6, Application PC/TUS9610602
XX
CC      Sequence 6, Application PC/TUS9610602
CC      GENERAL INFORMATION:
CC      APPLICANT: The General Hospital Corporation
CC      TITLE OF INVENTION: INHIBITION OF HEPATITIS B REPLICATION
CC      NUMBER OF SEQUENCES: 14
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE: Fish & Richardson P.C.
CC      STREET: 225 Franklin Street
CC      CITY: Boston
CC      STATE: MA
CC      COUNTRY: USA
CC      ZIP: 02110-2804
CC      COMPUTER READABLE FORM:
CC      MEDIUM TYPE: Floppy disk
CC      COMPUTER: IBM PC compatible
CC

```

CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US96/10602
CC FILING DATE:
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 60/017,814
CC FILING DATE: 20-JUN-1995
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Clark, Paul T.
CC REGISTRATION NUMBER: 30,162
CC REFERENCE/DOCKET NUMBER: 00786/282001
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 617/542-5070
CC TELEFAX: 617/542-8906
CC TELEX: 200154
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 397 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 397 AA; 45191 MW; 946620 CN;
SO
Query Match 3.5%; Score 99; DB 3; Length 397;
Best Local Similarity 48.7%; Pred. No. 3,69e+01;
Matches 19; Conservative 7; Mismatches 9; Indels 4; Gaps 3;
DB 152 RGRSPRRRT-PSPPR-R--RSGSPRRRSQSRLGPLVL 186
QY 6 RGRGPRSRGGRARARAGRCPRAROSPAPLIPDTVL 44
RESULT 11
ID PCT-US94-04496-30 STANDARD; PRT: 568 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 30, Application PC/TUS9404496
XX
CC Sequence 30, Application PC/TUS9404496
CC GENERAL INFORMATION:
CC APPLICANT: Croce, Carlo
CC APPLICANT: Canaan, Eli
CC TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
CC TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
CC TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1
CC NUMBER OF SEQUENCES: 86
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz 6
CC ADDRESSEE: Norris
CC STREET: One Liberty Place, 46th floor
CC CITY: Philadelphia
CC STATE: Pennsylvania
CC COUNTRY: USA
CC ZIP: 19103
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US94/04496
CC FILING DATE:
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Deluca Esq., Mark
CC REGISTRATION NUMBER: 33,229
CC REFERENCE/DOCKET NUMBER: TUD-1242

CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (215) 568-3100
CC TELEFAX: (215) 568-3439
CC INFORMATION FOR SEQ ID NO: 30:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 568 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 568 AA; 63367 MW; 1762870 CN;
SO
Query Match 3.5%; Score 99; DB 3; Length 568;
Best Local Similarity 26.7%; Pred. No. 3,69e+01;
Matches 32; Conservative 32; Mismatches 49; Indels 7; Gaps 7;
DB 132 RRLKAGCDPNRSIHSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSST 191
QY 102 RRLDPEGAAPVYVYSGKVOSSLNLIPDSSSLKLPSEDEADLTNGSSPSSEDDAL 161
DB 192 SFSKP-HK-LMKE-HKE-KPSKDSREHK-SAFKEPS-RDHNKSSESKPKPE-NKPLKE 244
QY 162 PGSPWRRKLRKCKEKKKEEPPDDDISPLPQPSRRNKRKHTALQRLREVNRKLD 221
RESULT 12
ID US-08-320-559-30 STANDARD; PRT: 568 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 30, Application US/08320559
XX
CC Sequence 30, Application US/08320559
CC Patent No. 5633135
CC GENERAL INFORMATION:
CC APPLICANT: Croce, Carlo
CC APPLICANT: Canaan, Eli
CC TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for
CC TITLE OF INVENTION: Detection and Treatment of Acute Leukemias
CC TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the
CC NUMBER OF SEQUENCES: 44
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5633135r1s
CC STREET: One Liberty Place - 46th floor
CC CITY: Philadelphia
CC STATE: PA
CC COUNTRY: USA
CC ZIP: 19103
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
CC COMPUTER: IBM PS/2
CC OPERATING SYSTEM: PC-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/320,559
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/062,443
CC FILING DATE: 14 MAY 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/07/971,094
CC FILING DATE: 30-OCT-92
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/07/888,830
CC FILING DATE: 27-MAY-92
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/07/805,093
CC FILING DATE: 11-DEC-91
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Deluca, Mark


```
CC REGISTRATION NUMBER: 33,229
CC REFERENCE/DOCKET NUMBER: T1U-0855
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (215) 568-3100
CC TELEFAX: (215) 568-3439
CC INFORMATION FOR SEQ ID NO: 30:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 568 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 568 AA; 63367 MW; 1762870 CN;
SO

Query Match 3.5%; Score 99; DB 1; Length 568;
Best Local Similarity 26.7%; Pred. No. 3,69e+01;
Matches 32; Conservative 32; Mismatches 49; Indels 7; Gaps 7;

Db 132 RRLKAGGDDPRSTHTSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 191
    ||||| : | : : : : : : : : : : : : : : : : : : : : : : :
Cc 102 RRLLDPGEPAPVPPYSGVQSSLNLPDNISSLLKICPEPEDEADLITGSSPSDDDL 161
    : : : : : | : | : | : : : : : : : : : : : : : : : : :
Cc 192 SFSEKP-HK-LMKE-HKE-KPSKDSREHK-SAFKEPS-RDHNNSSKESSKPKK-NKLPE 244
    : : : : : | : | : | : : : : : : : : : : : : : : : : :
Cc 162 PGSGWRRKKLRKCKEKKMEFPDQDISPLPQPSSRKRKRYTALQKLRVYNRLDD 221
Oy

RESULT 13 STANDARD; PRT; 30 AA.
ID US-08-218-608-10
AC xxxxxx
DT
XX
XX
XX
DE
XX
Sequence 10, Application US/08218608
CC Sequence 10, Application US/08218608
CC Patent No. 5607859
CC GENERAL INFORMATION:
CC APPLICANT: BIEMANN, KLAUS
CC APPLICANT: JUHASZ, PETER
CC TITLE OF INVENTION: METHODS AND PRODUCTS FOR MASS
CC TITLE OF INVENTION: SPECTROMETRIC MOLECULAR WEIGHT DETERMINATION OF POLYIONIC
CC TITLE OF INVENTION: ANALYTES EMPLOYING POLYIOTNIC REAGENTS
CC NUMBER OF SEQUENCES: 12
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
CC STREET: 600 ATLANTIC AVENUE
CC CITY: BOSTON
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02210
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/218,608
CC FILING DATE: 28-MAR-1994
CC CLASSIFICATION: 436
CC ATTORNEY/AGENT INFORMATION:
CC NAME: GATES, EDWARD R.
CC REGISTRATION NUMBER: 31,616
CC REFERENCE/DOCKET NUMBER: M0656/7013
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 617-720-3500
CC TELEFAX: 617-720-2441
CC INFORMATION FOR SEQ ID NO: 10:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 30 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
```

[illegible]

OY 64 VPVAPLPAKPEQSDSDSEGAAGPAGADPTLVRRRR 102

RESULT 15
ID US-08-240-717A-1 STANDARD: PRT; 123 AA.
AC xxxxxx
XX
DT
XX
DE
XX

Sequence 1, Application US/08240717A

CC Sequence 1, Application US/08240717A
CC Patent No. 5965353

CC GENERAL INFORMATION:
CC APPLICANT: Middelorp, Jaap M.
CC TITLE OF INVENTION: Epstein Barr Virus Peptides and
CC TITLE OF INVENTION: Antibodies Against These Peptides
CC NUMBER OF SEQUENCES: 6

CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Akzo No. 5965353e1 Patent Dept.

CC STREET: 1300 Piccard Drive, Suite 206
CC CITY: Rockville
CC STATE: Maryland
CC COUNTRY: US

CC ZIP: 20850

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patentin Release #1.0, Version #1.30

CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/240,717A

CC FILING DATE: 11-MAY-1994

CC CLASSIFICATION: 435

CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: EP 92.202797.4

CC FILING DATE: 14-SEP-1992

CC ATTORNEY/AGENT INFORMATION:
CC NAME: Gormley, Mary E.

CC REGISTRATION NUMBER: 34,409

CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (301) 258-5200

CC TELEFAX: (301) 977-0847

CC INFORMATION FOR SEQ ID NO: 1:

CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 123 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide

CC SEQUENCE 123 AA; 12990 MM; 70971 CN;

Query Match 3.4%; Score 97; DB 2; Length 123;

Best Local Similarity 28.3%; Pred. No. 4.85e+01;

Matches 28; Conservative 26; Mismatches 38; Indels 7; Gaps 7;

DB 21 RARG-RGRG-RGERRRPSSSSSSGSPRRPPGRPFPHYGEADYFEYHOGGPDG 78
QY 6 RGRGPRSGRGARARARCRPARASPARLPDT-VLVLDVSDSE-EVLEVADPVE 63
DB 79 EP-D-VPVGAIEGGPADDPGEGSTGPRGGDG-GRRRK 114
QY 64 VPVAPLPAKPEQSDSDSEGAAGPAGADPTLVRRRR 102

Search completed: Fri Feb 18 12:02:29 2000
Job time : 27 secs.

(TM)

Distribution rights by Oxford Molecular Ltd

ot generated

1 МАЕРІ

Gap 11

122810 seqs, 40068593 residues

Listing first 45 summaries

1:plr1 2:plr2 3:plr3 4:plr4

Mean 49.563; Variance 122.566; scale 0.404

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	161	5.6	95	2	JC4760	SM73 protein - human	2.27e+00
2	152	5.3	91	2	JC5582	SM73 protein - Caenor	2.96e+00
3	129	4.5	1402	2	LD6707	initiation factor 4-9	1.54e+00
4	125	4.4	101	2	S63999	SM73 protein - yeast	4.34e+00
5	117	4.1	771	2	B70564	hypothetical protein	3.25e+00
6	113	4.0	673	2	S60173	fragile X mental reta	8.63e+00
7	114	4.0	932	2	AA2632	cell adhesion molecul	6.77e+00
8	115	4.0	1870	2	S37671	bact2 protein - human	5.31e+00
9	115	4.0	1872	2	S36152	bact2 protein - human	5.31e+00
10	113	4.0	2142	2	B35098	MHC class III histoco	8.63e+00
11	111	3.9	359	1	U51734	transcription factor	1.40e+00
12	111	3.9	1050	2	U51034	gene P1 protein - fru	1.40e+00
13	111	3.9	229	1	WMAD51	late 33K protein - hu	2.24e+00
14	109	3.8	340	1	B55973	transcription factor	2.24e+00
15	109	3.8	380	2	T00792	hypothetical protein	2.24e+00
16	110	3.8	385	1	I79498	GTP cyclohydrolase II	1.77e+00
17	109	3.8	410	2	E70579	probable muc protein	2.24e+00
18	109	3.8	431	2	S09824	hypothetical protein	2.24e+00
19	110	3.8	456	2	S36513	I2 protein - human pa	1.77e+00
20	109	3.8	546	2	A45195	adenylyl cyclase type	2.24e+00
21	110	3.8	657	2	S05517	lamn - chicken	1.77e+00
22	109	3.8	1184	2	A42904	adenylyl cyclase type	2.24e+00
23	108	3.8	1210	2	A48001	phospholipase C (EC 3	2.84e+00

45	104	3.6	265	1	LPRA1B	apolipoprotein A-I pr	7.16e+00
44	106	3.7	3414	1	GNMVE	eye development prote	4.52e+00
43	105	3.7	1893	2	A56158	genome development	5.69e+00
42	106	3.7	1703	2	S13441	EF protein - streptoc	4.52e+00
41	105	3.7	1822	2	S13441	EF protein - yeast	4.52e+00
40	105	3.7	1234	2	I38894	phospholipase C beta	5.69e+00
39	107	3.7	1107	2	E65226	phospholipase C beta	5.69e+00
38	105	3.7	1051	2	S27002	phospholipase C (EC	5.69e+00
37	107	3.7	913	2	S20590	exo-alpha-1,4-lidase	4.52e+00
36	106	3.7	676	1	EDB823	immediate-early prote	4.52e+00
35	107	3.7	676	2	EDB822	immediate-early prote	3.58e+00
34	107	3.7	525	2	S36405	nucleocapsid protein	3.58e+00
33	105	3.7	358	1	A55973	transcription factor	5.69e+00
32	106	3.7	349	1	I50369	transcription factor	4.52e+00
31	106	3.7	349	1	I67417	transcription factor	4.52e+00
29	106	3.7	349	1	I53277	transcription factor	4.52e+00
28	105	3.7	338	2	I38587	retrovirus-related le	5.69e+00
27	110	3.8	2424	2	I46479	calcium channel BI-2	1.77e+00
26	110	3.8	2424	2	I46480	calcium channel BI-2	1.77e+00
25	108	3.8	2284	1	GNVSV	genome polypotein -	2.84e+00
24	110	3.8	1302	2	A41249	multidrug resistance	1.77e+00

ALIGNMENTS

```
ENTRY      1 Jc4760 #type complete  
TITLE     SMT3 protein - human  
ORGANISM  #formal_name Homo sapiens #common name man  
DATE      10-May-1996 #sequence_revision 16-Aug-1996 #text_change  
          17-Mar-1999
```

```
ACCESSIONS JC4760  
REFERENCE   Mannen, H.; Tseng, H.M.; Cho, C.; Li, S.S.L.  
#authors    Biochem. Biophys. Res. Commun. (1996) 222:178-180  
#journal    Cloning and expression of human homolog HSMT3 to yeast  
#title       suppressor of MIF2 mutations in a centromere protein gene  
            accession MU0D:96212932
```

```
#cross_references JC4760  
#accession       UCJ4760  
##molecule_type mRNA  
##residues        1-95 #label MAN  
##cross_references GB:L76416
```

```
GENERICS  
#gene             GDB:SMT3H2; HSMT3; SMT3B  
#map_position     4  
#cross_references GDB:5752862
```

```
CLASSIFICATION  #superfamily yeast SMT3 protein  
SUMMARY         #length 95 #molecular-weight 10871 #checksum 3774
```

```
Query Match               5.6%; Score 161; DB 2; Length 95;  
Best Local Similarity 32.9%; Pred.No. 2,2/e-06;
```

```
Matches 25; Conservative 19; Mismatches 31; Indels 1; Gaps 1.
```

Dd 12 TENDNHINIKVAGGDGS-VVOFKIRKTRPSLKLKKAYCERGSLSMROIREFPDGQPINET 70
 ||::| | : | :::::: || | | | : || | :
Oy 335 TETSQEALRLRQGAEKRQMELISLPSPGLVKYLMSHREAMGLSGHILSFDFDTKLSGR 394
 :
Db 71 DTPAQLEMEDEEDITDV 86
 :||:| : | | : |
Oy 395 ELPADLGIESDLLEY 410

```
RESULT      2 JC5582 #type complete  
ENTRY TITLE SMT3 protein - Caenorhabditis elegans  
ORGAISM #formal_name Caenorhabditis elegans  
DATE 23-Sep-1997 #sequence_revision 23-Sep-1997 #text_change  
           17-Mar-1999
```

```
ACCESIONS JC5582  
REFERENCE Choudhuri, B.K.; Li, S.S.L.  
#authors Blochem. Biophys. Res. Commun. (1997) 234:788-792  
#journal
```

[illegible]

```
#gene SGD: SMT3
##cross-references SGD:S0002918; MIPS:YDR510w
#map_position 4R
#CLASSIFICATION #superfamily yeast SMT3 protein
#length 101 #molecular-weight 11597 #checksum 8485
SUMMARY
```

RESULT	ENTRY	5
TITLE	B70564	#type complete
ORGANISM	hypothetical protein RV3649 - Mycobacterium tuberculosis	1
DATE	(strain H37Rv)	
	#formal_name Mycobacterium tuberculosis	
	17-Jul-1998 #sequence_revision 17-Jul-1998	
	17-Jul-1998	
ACCESSIONS	B70564	
REFERENCE	A70500	
#authors	Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Church	

```

**molecule_type 2771
**residues GB:293456; GB:LL123456; NID:g3261770; PID:e316541
**cross-references PID:g105045
**experimental_source strain H37Rv
GENETICS
#gene RV3649
#length 771 #molecular-weight 81409 #checksum 4348
SUMMARY

```

Query Match 4.1%; Score 117; DB 2; Length 771;
 Best Local Similarity 24.6%; Pred. No. 3.25e-01;
 Matches 42: Conservative 47; Mismatches 72; Indels 10; Gaps 10;

Db 259 RTVALMEPARLSDVIG-EGHAPVRSAGAARVMADLIVEGAQTLTFVRSRRRAETAL 317
 213 REVNRKRLQDLRFCLSPKQHPALOSTDDEVVLEGPVLPQSSRLFTL-KIRCRADLVRL 271

Db 318 GARARLVDIAPLSDTVASVYRAGLYLAEDRSALHQAELGKATNLMLGVDIAGLD 377
 272 -P-VRMSEPLQWYVDMAN-HLG-VSPNR-IL-LFGESELSPTAPSTLKGVADI-ID 324

Db 378 AVVIAGFPGTVASFMOQAGRS-GRGOGALVYLIARDPLDFTLVHHPAL 427
 325 CVYLAASSSATETSOELRLRVGKKEKHWLEISLSPDPLKVMHYEEM 375

RESULT 6
 ENTRY 560173 #type complete
 TITLE fragile x mental retardation syndrome related protein FXR2 -
 ORGANISM human
 #formal_name Homo sapiens #common_name man
 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change
 15-Jan-1999

ACCESSIONS
 REFERENCE S60173
 #authors Zhang, Y.; O'Connor, J.P.; Stomil, M.C.; Stinivasan, S.;
 Dutra, A.; Nussbaum, R.L.; Dreyfuss, G.
 #journal EMBO J. (1995) 14:5358-5366
 The fragile x mental retardation syndrome protein interacts
 with novel homologs FXR1 and FXR2.
 #cross-references MVID:96080171
 #accession S60173
 #status preliminary
 #molecule_type mRNA
 #residues 1-673 #label ZHA
 #cross-references EMBL:U01501; NID:g1098636; PID:g1098637
 the authors translated the codon CTA for residue 59 as
 Thr, CTG for residue 91 as Thr, CTG for residue 164 as
 Thr, and CTC for residue 171 as Thr

GENETICS
 #gene GDB:FXR2
 ##cross-references GDB:1313721
 CLASSIFICATION #superfamily fragile x mental retardation syndrome protein
 KEYWORDS nucleolus; RNA binding
 FEATURE 232-261
 SUMMARY #domain ribonucleoprotein K similarity #label KH1
 #length 673 #molecular-weight 74128 #checksum 2966

Query Match 4.0%; Score 113; DB 2; Length 673;
 Best Local Similarity 32.4%; Pred. No. 8.63e-01;
 Matches 23: Conservative 18; Mismatches 26; Indels 4; Gaps 4;

Db 471 RDPPTGESSRRPTGRCGPPAPRPTSRVSSISSVLKDPDNPSTLDTSP-EP 529
 8 RPPRSRGGRARRAGARCR-CPRARQSPARLIPDTLVLDV-SDSD-EEVLEADPVEY 64

Db 530 PVDSEPEPPP 540
 65 PVARLPAPAKP 75

RESULT 7
 ENTRY A42632 #type complete
 TITLE cell adhesion molecule apCAM (clone d19) - California sea
 ORGANISM hare
 #formal_name Aplysia californica #common_name California sea
 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change
 28-Apr-1995
 ACCESSIONS A42632
 REFERENCE A42632

#authors Mayford, M.; Barzilai, A.; Keller, F.; Schacher, S.; Kandel,
 E.R.
 #journal Science (1992) 256:638-644
 #title Modulation of an NCAM-related adhesion molecule with
 long-term synaptic plasticity in Aplysia.
 #cross-references MVID:92263095
 #accession A42632
 #status preliminary; not compared with conceptual translation
 #molecule_type nucleic acid
 #residues 1-932 #label MAY
 #experimental_source CNS
 #note sequence extracted from NCBI backbone (NCBI:101342)
 #length 932 #molecular-weight 101521 #checksum 807

Query Match 4.0%; Score 114; DB 2; Length 932;
 Best Local Similarity 29.5%; Pred. No. 6.77e-01;
 Matches 28: Conservative 26; Mismatches 33; Indels 8; Gaps 8;

Db 826 RGRIGKGGEERKARD-GKDPAEDEKLKEDSVYENVEENKPDVYE-EQPEFER 883
 11 RSR-GGRARRARAGRCRPARQSPARLI-PDTLVLD-VSDSD-EEVLEADPVEY-P 65

Db 884 ETKPERTAPATGPAEADAKPEDLPAPAPTEIK 918
 66 VARLPAPAKPEDSDSDSEGAAG-PAGAPRTLYR 99

RESULT 8
 ENTRY S37671 #type complete
 TITLE bat2 protein - human
 ORGANISM #formal_name Homo sapiens #common_name man
 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
 10-Sep-1997

ACCESSIONS
 REFERENCE S37671
 #authors Bougueleret, L.
 #submission submitted to the EMBL Data Library, August 1992
 #accession S37671
 #status preliminary
 #molecule_type DNA
 #residues 1-1870 #label BOU
 #cross-references EMBL:Z15025; NID:g29374; PID:g29375

GENETICS
 #map_position 6p21.3
 #introns 38/2; 97/2; 129/3; 154/1; 202/1; 252/3; 279/2; 327/1; 357/2;
 429/3; 588/1; 651/1; 751/1; 821/2; 1836/2; 1523/1; 1368/3;
 1612/1; 1640/3; 1715/3; 1762/3; 1842/3
 SUMMARY #length 1870 #molecular-weight 199223 #checksum 7735

Query Match 4.0%; Score 115; DB 2; Length 1870;
 Best Local Similarity 29.5%; Pred. No. 5.31e-01;
 Matches 31: Conservative 31; Mismatches 37; Indels 6; Gaps 6;

Db 1023 RGRGEYFANG-RGRGTGGRGGAQANSVTESEFEMMGVE-VGQGDQTLTLLPAA-L 1079
 6 RGRGP-RSRGGRARRARAGRCRPARQSPARLIPDTLVLDVSDSDDEVLEADPVEY 64

Db 1080 P-ARRGARVQSMKSPSGAGSGAQKQAPRMYITWLQTRRLP 1123
 65 PVARLPAPAKPEDSDSDSEGAAGPAGAPRTLYR-RRRLTLDP 108

RESULT 9
 ENTRY S36152 #type complete
 TITLE bat2 protein - human
 ORGANISM #formal_name Homo sapiens #common_name man
 06-Jun-1995 #sequence_revision 17-Nov-1995 #text_change
 17-Nov-1995
 ACCESSIONS S36152
 REFERENCE S36152
 #authors Irls, F.J.M.; Bougueleret, L.; Priour, S.; Caterina, D.;
 Primas, G.; Perrot, V.; Jurka, J.; Rodriguez-Tome, P.;
 Claverie, J.M.; Dausset, J.; Cohen, D.

```

#journal      Nature Genet. 1993;3:137-145
#title        Dense Alu clustering and a potential new member of the
              NFkapab family within a 90 Kilobase HLA Class III segment.
#accession    S36152
##status      preliminary
##molecule_type DNA
##residues     1-1872 ##label IRI
##cross-references EMBL:Z15025
##note         In the authors' translation residues 32-34 are shown
               after residue 4 and, consequently, residues 5-31 are
               displaced three codons to the right
               the authors translated the codon AAT for residue 1000 as
               His
GENETICS
#introns
38/2; 97/2; 129/3; 154/1; 202/1; 252/3; 279/2; 327/1; 357/2;
429/3; 588/1; 651/1; 751/1; 821/2; 1436/2; 1555/1; 1568/3;
1612/1; 1640/3; 1715/3; 1762/3; 1786/3; 1842/3
length 1872 #molecular-weight 199338 #checksum 7324
MMARY
Query Match          4.0%; Score 115; DB 2; Length 1872;
Best Local Similarity 29.5%; Pred. No. 5.31e+01;
Matches 31; Conservative 31; Mismatches 37; Indels 6; Gaps 6;
Db 1024 RGRGEYFARG-RGFRGYGGRGGAQANSVTESEFEEMNGVE-VGGGDQTLLPEAA-L 1080
||||| |||| | : |||| | : : : : : : : : : : : : : : : :
Oy 6 RGRP-KSRGRGARRRRAGRCPRKRSPPALLIDTVLVLDVSDEVLVDADPVEV 64
||| : : : : : : : : : : : : : : : : : : : : : : :
DB 1081 P-ARHGAVOSMRKSPSGAGSQAOKAARPMRYIMLTQRRLPH 1124
||| : : : : : : : : : : : : : : : : : : : : : : :
Oy 65 PVARKLPAPAKPEQSDSDSEGAEAGPAAPRTLYRR-RRRRLDP 108
||| : : : : : : : : : : : : : : : : : : : : : : :
RESULT 10
ENTRY B35098 #type complete
TITLE MHC class III histocompatibility antigen HLA-B-associated
TRANSCRIPT 2 - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 10-Aug-1990 #sequence_revision 06-Nov-1992 #text_change
10-Sep-1997
ACCESSIONS B35098
REFERENCE A35098
#authors Banerji, J.; Sands, J.; Strominger, J.L.; Spies, T.
#journal Proc. Natl. Acad. Sci. U.S.A. 1990; 87:2374-2378
#title A gene pair from the human major histocompatibility complex
       encodes large proline-rich proteins with multiple repeated
       motifs and a single ubiquitin-like domain.
#cross-references MUID:90192810
#accession B35098
#status preliminary
##molecule_type mRNA
##residues 1-2142 ##label BAN
##cross-references GB:M3509; NID:g179338; PID:g179339; GB:M31293
##note the authors translated the codon AGT for residue 97 as
       Gly
SUMMARY #length 2142 #molecular-weight 227840 #checksum 5704
Query Match          4.0%; Score 113; DB 2; Length 2142;
Best Local Similarity 29.5%; Pred. No. 8.63e+01;
Matches 31; Conservative 30; Mismatches 38; Indels 6; Gaps 6;
Db 1012 RGREYFARG-RGFRGYGGRGGAQANSVTSEFEEMNGVE-VGGGDQTLLPEAA-M 1068
||||| |||| | : |||| | : : : : : : : : : : : : : : : :
Oy 6 RGRP-KSRGRGARRRRAGRCPRKRSPPALLIDTVLVLDVSDEVLVDADPVEV 64
||| : : : : : : : : : : : : : : : : : : : : : : :
Db 1069 P-ARHGAVOSMRKSPSGAGSQAOKAARPMRYIMLTQRRLPH 1112
||||| |||| | : |||| | : : : : : : : : : : : : : : : :
Oy 65 PVARKLPAPAKPEQSDSDSEGAEAGPAAPRTLYRR-RRRRLDP 108
||| : : : : : : : : : : : : : : : : : : : : : : :
RESULT 11
ENTRY I51734 #type complete
TITLE transcription factor Isl-2 - zebra fish
ALTERNATE_NAMES insulin enhancer-binding protein Isl-2; Islc-2 protein
```

```
ORGANISM      #formal_name Brachydanio rerio #common_name zebra fish  
DATE          13-Mar-1997 #sequence_revision 13-Mar-1997 #text_change  
ACCESSIONS    05-Sep-1997  
REFERENCE     151734; S57407  
#authors      Tokumoto, M.; Gong, Z.; Tsubokawa, T.; Hew, C.L.; Uyemura,  
              K.; Hotta, Y.; Okamoto, H.  
#title        Dev. Biol., (1995) 171:578-589  
              Molecular heterogeneity among primary motoneurons and within  
              myotomes revealed by the differential mRNA expression of  
              novel Islet-1 homologs in embryonic zebrafish.  
#cross-references MUID:96005022  
#accession    151734 preliminary; translated from GB/EMBL/DDBJ  
##status      ##molecule_type mRNA  
##residues    1-359 ##label TOK  
#cross-refe-nces GB:D38453; NID:g1037165; PID:g1037166  
REFERENCE     S57407  
#authors      Appel, B.; Korzh, V.; Glasgow, E.; Thor, S.; Edlund, T.;  
              David, L.; Eisen, J.  
#submission   Submitted to the EMBL Data Library, June 1995  
#description   Motoneuron fate specification and patterned LIM homeobox gene  
               expression in embryonic zebrafish.  
#accession    S57407 preliminary  
##status      ##molecule_type mRNA  
##residues    1-359 ##label APP  
#cross-refe-rences EMBL:X88805; NID:g871000; PID:g871001  
GENETICS      #gene         Isl-2  
CLASSIFICATION superfamily transcription factor Isl-1; homeobox homology;  
KEYWORDS       LIM metal-binding repeat homology  
                DNA binding; duplication; homeobox; nucleus; transcrip-tion  
                regulation; zinc  
  
FEATURE  
27-80          #domain LIM metal-binding repeat homology #label LIM1\  
89-142         #domain LIM metal-binding repeat homology #label LIM2\  
192-248        #domain homeobox homology #label HOX  
SUMMARY        #length 359 #molecular_weight 40227 #checksum 7994  
  
Query Match    3.9%; Score 111; DB 1; Length 359;  
Best Local Similarity 30.3%; Pred. No. 1,40e+00;  
Matches 23; Conservative 23; Mismatches 24; Indels 6; Gaps 6;
```

ID	Type	Description
Dbj	LIM1	PYROPFHRNHHKQSEKTTRVTVLENEKLHRTTCYNAPRPD-ALM-KE-QLVEXTG- 229
Oj	LIM2	PLPQPSSRRKSRSKHTEAOKLRVN-N-KRIODLSRCLSPKHOHPAQSTGTDEVLVEGP 249
Dbl	Hox	LSPRVIRWFQNKRCK 245
Oj	Hox	VLPQSRLFTLKIRCR 265

Title	Accessions	Date	#Authors	Status	Molecule Type	Label
Sil1034	SI1034	21-Nov-1993	J.M.Oliver	Final	mRNA	TOK
Sil1034	SI1034	24-Sep-1998	J.A.Lepeant	Preliminary	mRNA	APP
Sil1034	SI1034	21-Dec-1995	F.Dubertret, M.L.Therond, P.Clavetle, J.M.	Translated	mRNA	TOK

Title	Accessions	Date	#Cross-References	Status	Molecule Type	Label
Sil1034	SI1034	21-Dec-1995	G.B.D38453, N.G.1037165, P.ID.g1037166	Translated	mRNA	TOK
Sil1034	SI1034	24-Sep-1998	B.Appel, V.Korzh, E.Glasgow, S.Thor, S.Edlund, T.David, L.Eisen	Submitted	mRNA	APP

Title	Accessions	Date	#Cross-References	Status	Molecule Type	Label
Sil1034	SI1034	21-Dec-1995	E.MBL.X88805, N.ID.g871000, P.ID.g871001	Translated	mRNA	APP

```

#gene FlyBase:FBp1
#cross-references FlyBase:FBgn0000639
#introns 31/1
SUMMARY length 1030 #molecular-weight 119364 #checksum 1902

Query Match 3.9%; Score 111; DB 2; Length 1030;
Best Local Similarity 20.9%; Pred. No. 1.40e+00;
Matches 42; Conservative 66; Mismatches 76; Indels 17; Gaps 15;

Db 559 TDDLELHMQRRQRLQKH-QNDDDD-DD-NDDVNVYHROGLSKSRSLPMLRQONNL 615
Oy 156 SEDDLPLPS-GSPWRRKKLRKCKEKKMEFPDQIDISLPQSSSRKSKRTAQLKLE 214
Db 616 SEIVAHNRQ-LVARLNQESIAGOLIEEOQLINPRLQSEYALRLNQRINSORR 674
Oy 215 VKRRQDLRSCSPKQHQSPALQS--TDDEVYLVESPVLPQSSRL-FTLK-IRCRADLVR 270
Db 675 OVLAIQIGIQRIGIYIGVLSQVNVNLSRGVIDQROVESIADVLGRLGOVGIMTII 734
Oy 271 --LP-V-RMSEPLQNVVDHMANHGLVSPNRIILFG-ESE-LSPTPTSTL-KLGVDIIT 323
Oy 735 ROYVDNNSQIDRNGLGIRL 755
Oy 324 DCVVLASSEATE-TSQELRL 343

RESULT 13
ENTRY WMA051 #type complete
TITLE late 33k protein - human adenovirus 5
ORGANISM #formal_name Mastadenovirus h5 #common_name human adenovirus
#note host Homo sapiens (man)
DATE 30-Sep-1992 #sequence-revision 30-Sep-1992 #text-change 04-Mar-1994

ACCESSIONS F39449
REFERENCE A39449
#authors Chroboczek, J.; Bieber, F.; Jacrot, B.
#journal Virology (1992) 186:280-285
#title The sequence of the genome of adenovirus type 5 and its comparison with the genome of adenovirus type 2.
#cross-references MUID:92087470
#accession F39449
#molecule_type DNA
#residues 1-229 #label CHR
#cross-references GB:M73260
CLASSIFICATION #superfamily adenovirus late 33k protein
KEYWORDS late protein
SUMMARY #length 229 #molecular-weight 25164 #checksum 9588

Query Match 3.8%; Score 108; DB 1; Length 229;
Best Local Similarity 31.6%; Pred. No. 2.84e+00;
Matches 18; Conservative 14; Mismatches 24; Indels 1; Gaps 1;

Db 42 IDEASEVEVSDETPSPVAPSPA-POKSATGSGMATSPAOAPPALVRRPNRR 97
Oy 48 VSDSEVLEVDPEVPAARLPAAPAKPEQSDSDSEGAABEPAGAPRLVRRRRR 104

RESULT 14
ENTRY B55973 #type fragment
TITLE transcription factor Isl-2b - chinook salmon (fragment)
ALTERNATE_NAMES insulin enhancer-binding protein Isl-2b; Isllet-2b protein
ORGANISM #formal_name Oncorhynchus tshawytscha #common_name chinook salmon
DATE 03-Oct-1995 #sequence-revision 03-Oct-1995 #text-change 05-Sep-1997
ACCESSIONS B55973; S52091; S19956
REFERENCE A55973
#authors Gong, Z.; Hui, C.; Hew, C.L.
#journal J. Biol. Chem. (1995) 270:3335-3345
#title Presence of Isl-1-related LIM domain homeobox genes in teleost and their similar patterns of expression in brain and spinal cord.

```

```

#accession B55973
#molecule_type mRNA
#residues 1-340 #label GON
#cross-references EMBL:X64884; NID:g64208; PID:g64209
#note authors translated the codon CTG for residue 251 as Met, and ATG for residue 256 as Leu

REFERENCE S52089
#authors Gong, Z.; Hew, C.L.
#journal Biochim. Biophys. Acta (1995) 1260:349-354
#title Several splicing variants of Isl-1 like genes in the chinook salmon (Oncorhynchus tshawytscha) encode truncated transcription factors containing a complete LIM domain.
#cross-references MUID:95178560
#accession S52091
#status preliminary
#molecule_type mRNA
#residues 1-340 #label G02
#cross-references EMBL:X64884; NID:g64208; PID:g64209
#note the authors translated the codon CAG for residue 105 as His, GTG for residue 107 as Leu, CTG for residue 251 as Met, and ATG for residue 256 as Leu

GENETICS
#gene Isl-2b
CLASSIFICATION #superfamily transcription factor Isl-1; homeobox homology; LIM metal-binding repeat homology
KEYWORDS DNA binding; duplication; homeobox; nucleus; transcription regulation; zinc
FEATURE 9-62
7-124
173-229
SUMMARY #length 340 #checksum 5243

Query Match 3.8%; Score 109; DB 1; Length 340;
Best Local Similarity 28.9%; Pred. No. 2.24e+00;
Matches 24; Conservative 24; Mismatches 29; Indels 6; Gaps 6;

Db 148 FHIADEVSVROPPIHNVKROSEKTRVRYLNEKQLHTRFCYANPNRPD-ALM-XE-Q 204
Oy 185 FPDQDISPLPQPSNRKSKRTKTEALQKREV-N-KRLQDLRSCLSKQHQSPALQSTDE 242
Db 205 LVEMTG-LSPRVIRYWFQNRCK 226
Oy 243 VLVGEGPVLPQSSRLFTLKRCR 265

RESULT 15
ENTRY T00792 #type complete
TITLE hypothetical protein F24L7.10 - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear cress
DATE 12-Feb-1999 #sequence-revision 12-Feb-1999 #text-change 12-Feb-1999

ACCESSIONS T00792
REFERENCE Z14204
#authors Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.
#submission submitted to the EMBL Data Library, February 1998
#description Arabidopsis thaliana chromosome II BAC F24L7 genomic sequence.
#accession T00792
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-380 #label ROU
#cross-references EMBL:AC003974; NID:g2914688; PID:g2914697
#experimental_source cultivar Columbia

GENETICS
#map_position 2
#introns 55/3; 116/1; 161/3; 237/3; 287/3; 357/3
#note F24L7.10
SUMMARY #length 380 #molecular-weight 42844 #checksum 4620

```

 W P O S E H
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Msrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Fri Feb 18 11:58:36 2000; MasPar time 13.08 Seconds
 890.539 Million cell updates/sec
 Modular output not generated.

Title: >US-08-755-584-2
 Description: (1-412) from 5858711.pep
 Perfect Score: 2858
 Sequence: 1 MAEPDNGRGRNGRNGR.....GKELPADLGESDLEWVG 412

Scoring table:
 PAM 150
 Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: swiss-prot37
 1:swissprot

Statistics: Mean 50.879; Variance 107.959; scale 0.471

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.	
1	161	5.6	95	1	SM32_HUMAN	UBIQUITIN-LIKE PROTEIN	8.16e-08
2	159	5.6	100	1	SMT3_ORYSA	UBIQUITIN-LIKE PROTEIN	1.58e-07
3	160	5.6	104	1	SMT3_ARATH	UBIQUITIN-LIKE PROTEIN	1.14e-07
4	152	5.3	91	1	SMT3_CAEEL	UBIQUITIN-LIKE PROTEIN	1.55e-06
5	149	5.2	101	1	SMT3_HUMAN	UBIQUITIN-LIKE PROTEIN	4.07e-06
6	141	4.9	103	1	SM31_HUMAN	UBIQUITIN-LIKE PROTEIN	5.10e-05
7	129	4.5	1402	1	IF4G_RABIT	EUKARYOTIC TRANSLATION	1.98e-03
8	126	4.4	90	1	SMT3_SCHPO	UBIQUITIN-LIKE PROTEIN	4.80e-03
9	125	4.4	101	1	SMT3_YEAST	UBIQUITIN-LIKE PROTEIN	6.43e-03
10	113	4.0	673	1	FXR2_HUMAN	FRAGILE X MENTAL RETAR	1.93e-01
11	113	4.0	2142	1	BAT2_HUMAN	LARGE PROLINE-RICH PRO	1.93e-01
12	111	3.9	359	1	ISL2_BRARE	INSULIN GENE ENHANCER	3.33e-01
13	112	3.9	454	1	VE2_HPV37	REGULATORY PROTEIN E2	2.33e-01
14	112	3.9	707	1	RHO_STRLI	TRANSCRIPTION TERMINAT	2.53e-01
15	111	3.9	1030	1	FBP1_DROME	FAT-BODY PROTEIN-1 PRE	3.33e-01
16	108	3.8	229	1	V33P_ADE05	33 KD PHOSPHOPROTEIN	7.43e-01
17	110	3.8	267	1	RS2_DROME	40S RIBOSOMAL PROTEIN	4.35e-01
18	109	3.8	340	1	IS2B_ONCTS	INSULIN GENE ENHANCER	5.69e-01
19	110	3.8	385	1	GCH2_AOBR	GTP CYCLOHYDROLASE II	4.35e-01
20	109	3.8	431	1	UL61_HCMVA	HYPOTHETICAL PROTEIN U	5.69e-01
21	110	3.8	476	1	VL2_HPV32	MINOR CAPSID PROTEIN L	4.35e-01
22	109	3.8	619	1	BCHD_CHLVI	MAGNESIUM-CHELATASE 67	5.69e-01
23	110	3.8	657	1	LAMA_CHICK	LAMIN A	4.35e-01

ID	Score	Query Match	Length	ID	Description	Pred. No.	
24	109	3.8	708	1	MR11_HUMAN	MRE11 HOMOLOG.	5.69e-01
25	109	3.8	1184	1	CV45_CANFA	ADENYLATE CYCLASE, TYP	5.69e-01
26	110	3.8	1302	1	MDR4_DROME	MULTIDRUG RESISTANCE P	4.35e-01
27	108	3.8	2284	1	POLI_GFLV	RNAI POLYPROTEIN (253	7.43e-01
28	110	3.8	2424	1	CB12_RABIT	BRAIN CALCIUM CHANNEL	4.35e-01
29	107	3.7	272	1	RS2_CAEEL	PROBABLE 40S RIBOSOMAL	1.26e+00
30	106	3.7	349	1	ISL1_CHICK	INSULIN GENE ENHANCER	1.26e+00
31	106	3.7	349	1	ISL1_HUMAN	INSULIN GENE ENHANCER	1.26e+00
32	105	3.7	358	1	IS2A_ONCTS	INSULIN GENE ENHANCER	1.63e+00
33	107	3.7	525	1	NCAP_PPV	NUCLEOCAPSID PROTEIN	9.68e-01
34	107	3.7	676	1	ICPO_HSVB	TRANS-ACTING TRANSCRIP	1.26e+00
35	106	3.7	676	1	ICPO_HSVB	TRANS-ACTING TRANSCRIP	1.26e+00
36	107	3.7	1107	1	YBP_ECOLI	HYPOTHETICAL 123.8 KD	9.68e-01
37	105	3.7	1234	1	PIB3_HUMAN	1-PHOSPHATIDYLINOSITOL	1.63e+00
38	106	3.7	1703	1	SNR2_YEAST	TRANSCRIPTION REGULATO	1.26e+00
39	107	3.7	3210	1	CENF_HUMAN	CENP-F KINETOCHORE PRO	9.68e-01
40	106	3.7	3414	1	POLG_TBEV	GENOME POLYPROTEIN (CO	1.26e+00
41	104	3.6	140	1	YV16_HUMAN	HYPOTHETICAL 14 KD PRO	2.12e+00
42	103	3.6	496	1	YA31_SCHPO	HYPOTHETICAL 57.8 KD P	2.74e+00
43	104	3.6	1679	1	YIO9_YEAST	HYPOTHETICAL 195.1 KD	3.12e+00
44	102	3.6	1733	1	VNDA_PPVKA	PROBABLE NUCLEAR ANTIG	3.35e+00
45	103	3.6	2663	1	CENF_HUMAN	CENTROMERIC PROTEIN E	2.74e+00

ALIGNMENTS

RESULT 1
 ID SM32_HUMAN STANDARD: PRT: 95 AA.
 AC P55855;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE UBIQUITIN-LIKE PROTEIN SMT3B (SENTRIN 2).
 GN SMT3B2 OR SMT3B.
 OS HOMO SAPIENS (HUMAN), AND BOS TAURUS (BOVINE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=HUMAN; TISSUE=BRAIN;
 RX MEDLINE: 97237059.
 RA LAPENTA V., CHIBRAZZI P., VAN DER SPEK P.J., PIZZUTI A.,
 RA HANOKA F., BRAHE C.;
 RT "SMT3A, a human homologue of the S. cerevisiae Smt3 gene, maps to
 RT chromosome 21qter and defines a novel gene family";
 RL GENOMICS 40:362-367(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=BOVINE; TISSUE=ENDOMETRIUM;
 RA O'F., BERRY E.S.;
 RA SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 -1- SIMILARITY: TO UBIQUITIN. BELONGS TO THE SMT3 SUBFAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 or send an email to license@sdb-sdb.ch).

 CC EMBL: X95985; E274602; -;
 DR EMBL: U89439; G1888536; -;
 DR MIM: 603042; -;
 DR PROSITE: P550053; UBIQUITIN_2; 1.
 FT DOMAIN 16 95 UBIQUITIN-LIKE.
 SO SEQUENCE 95 AA: 10871 MW: E49P30B9 CRC32:
 Query Match 5.6%; Score 161; DB 1; Length 95;
 Best Local Similarity 32.9%; Pred. No. 8.16e-08;
 Matches 25; Conservative 19; Mismatches 31; Indels 1; Gaps 1;
 DB 12 TENNHNINKVAGGOGS-VVQFKIKRHPTLSKIMKAYCERGGISNMQIFRFDGQINFT 70

```

QY 335 TETSOELRLVQGEKHEKHMLETISPDSPKLVMSHYEANGLSGKLSFFDGTKLSGK 394
DB 71 DTPAOLEMEDDITDY 86
QY 395 ELPADLGLESGLIEV 410

RESULT 2
ID SMT3_ORYSA STANDARD; PRT: 100 AA.
AC P55853;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE UBIQUITIN-LIKE PROTEIN SMT3.
GN SMT3.
OS ORYZA SATIVA (RICE).
OC EUKARYOTA; VIRIDIPHYTES; STREPTOPHYTA; EMAROPHYTA; TRACHEOPHYTA;
EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; POALES;
OC PACEAE; ORYZA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97237059.
RA LAPENTA V., CHIRAZZI P., VAN DER SPEK P.J., PIZZUTI A.,
RA HANAOKA F., BRAHE C.;
RT "SMT3A, a human homologue of the S. cerevisiae SMT3 gene, maps to
RT chromosome 21qter and defines a novel gene family.";
RL GENOMICS 40:362-367(1997).
CC -1- SIMILARITY: TO UBIQUITIN. BELONGS TO THE SMT3 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X99608; E259739;
DR PROSITE: PSS0053; UBIQUITIN_2; 1.
FT DOMAIN 19 96 UBIQUITIN-LIKE.
SQ SEQUENCE 100 AA; 10928 MW; 4EFFDFD CRC32;

Query Match
Best Local Similarity 29.0%; Pred. NO. 1.58e-07;
Matches 20; Conservative 23; Mismatches 25; Indels 1; Gaps 1;

DB 21 INLVKGGDGNVFF-FRIKRSQTKLMNAYCDQSVNMNIAFLDGRRLRGEDTPDEL 79
QY 341 LRLVQGEKHEKHMLETISPDSPKLVMSHYEANGLSGKLSFFDGTKLSGKELPADL 400

RESULT 3
ID SMT3_ARATH STANDARD; PRT: 104 AA.
AC P55853;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE UBIQUITIN-LIKE PROTEIN SMT3.
GN SMT3.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPHYTES; STREPTOPHYTA; EMAROPHYTA; TRACHEOPHYTA;
EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97237059.
RA LAPENTA V., CHIRAZZI P., VAN DER SPEK P.J., PIZZUTI A.,
RA HANAOKA F., BRAHE C.;

```

```

RT "SMT3A, a human homologue of the S. cerevisiae SMT3 gene, maps to
RT chromosome 21qter and defines a novel gene family.";
RL GENOMICS 40:362-367(1997).
CC -1- SIMILARITY: TO UBIQUITIN. BELONGS TO THE SMT3 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X99609; E259962;
DR PROSITE: PSS0053; UBIQUITIN_2; 1.
FT DOMAIN 16 93 UBIQUITIN-LIKE.
SQ SEQUENCE 104 AA; 11654 MW; CD706439 CRC32;

Query Match
Best Local Similarity 25.7%; Pred. NO. 1.14e-07;
Matches 19; Conservative 26; Mismatches 28; Indels 1; Gaps 1;

DB 13 DCGAHINLVKGGDGNVFF-FRIKRSQTKLMNAYCDQSVNMNIAFLDGRRLRAEQ 71
QY 336 ETSOELRLVQGEKHEKHMLETISPDSPKLVMSHYEANGLSGKLSFFDGTKLSGK 395
DB 72 TPDELDMEDDITDY 85
QY 396 LPADLGLESGLIE 409

RESULT 4
ID SMT3_CAEEL STANDARD; PRT: 91 AA.
AC P55853;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE UBIQUITIN-LIKE PROTEIN SMT3.
GN SMT3.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97237059.
RA LAPENTA V., CHIRAZZI P., VAN DER SPEK P.J., PIZZUTI A.,
RA HANAOKA F., BRAHE C.;
RT "SMT3A, a human homologue of the S. cerevisiae SMT3 gene, maps to
RT chromosome 21qter and defines a novel gene family.";
RL GENOMICS 40:362-367(1997).
GN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97318858.
RA CHOUHURY B.K., LI S.S.;
RT "Identification and characterization of the SMT3 cDNA and gene from
RT nematode Caenorhabditis elegans.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 234:788-791(1997).
CC -1- SIMILARITY: TO UBIQUITIN. BELONGS TO THE SMT3 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X99600; E259960;
DR EMBL: U94830; G2341095;
DR PROSITE: PSS0053; UBIQUITIN_2; 1.
FT DOMAIN 13 91 UBIQUITIN-LIKE.
SQ SEQUENCE 91 AA; 10222 MW; F8E538FD CRC32;

```

Query Match 5.3%; Score 152; DB 1; Length 91;
 Best Local Similarity 22.5%; Pred. No. 1.55e-06;
 Matches 18; Conservative 31; Mismatches 30; Indels 1; Gaps 1;

DB 5 AAQAGDMEYIKYKVVGDSDNEV-HFRVYKGTSMALKKSVADRTGVAVNSIRFLFDGR 63
 QY 331 SEAEETSGEQLRLRVQGEKHEHMLEISLSPDSPKLKMSHYEAMGLSGHKLSFFDGT 390
 DB 64 INDDPTKLEMEDDVEV 83
 QY 391 LSGKELPADLGLSGDLLEV 410

RESULT 5
 ID SM31_HUMAN STANDARD; PRT; 101 AA.
 AC Q93068; P55856;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE UBIQUITIN-LIKE PROTEIN SMT3C (UBIQUITIN-HOMOLOGY DOMAIN PROTEIN PIC1)
 (UBIQUITIN-LIKE PROTEIN UBL1) (UBIQUITIN-RELATED PROTEIN SUMO-1)
 (GAP MODIFYING PROTEIN 1) (GmP1) (SENTRIN).
 SMT3H OR SMT3C.
 OS HOMO SAPIENS (HUMAN), AND MUS MUSCULUS (MOUSE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 CC PRIMATES; CATAARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES-HUMAN; TISSUE-BRAIN;
 RX MEDLINE; 97237059.
 RA LAPEENTA V., CHIDRAZZI P., VAN DER SPEK P.J., PIZZUTI A.,
 RA HANAOKA F., BRAHE C.;
 RT "SMT3A, a human homologue of the S. cerevisiae SMT3 gene, maps to
 RT chromosome 21qter and defines a novel gene family.";
 RL GENOMICS 40:362-367(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES-HUMAN; TISSUE-PLACENTA;
 RX MEDLINE; 96400311.
 RA BODDY M.N., HOWE K., ETKIN L.D., SOLOMON E., FREEMONT P.S.;
 RT "PIC 1, a novel ubiquitin-like protein which interacts with the pML
 RT component of a multiprotein complex that is disrupted in acute
 RT promyelocytic leukaemia.";
 RL ONCOGENE 13:971-982(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES-HUMAN;
 RX MEDLINE; 96411684.
 RA SHEN Z., PARDINGTON-PURTYMUN P.E., COMEAUX J.C., MOYZIS R.K.,
 RA CHEN D.J.;
 RT "UBL1, a human ubiquitin-like protein associating with human
 RT RAD51/RAD52 proteins.";
 RL GENOMICS 36:271-279(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC SPECIES-HUMAN;
 RX MEDLINE; 97148692.
 RA MAHAJAN R., DELPHIN C., GUAN T., GERACE L., MELCHIOR F.;
 RT "A small ubiquitin-related polypeptide involved in targeting RanGAP1
 RT to nuclear pore complex protein RanBP2.";
 RL CELL 88:97-107(1997).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC SPECIES-HUMAN;
 RX MEDLINE; 97133418.
 RA MATUNIS M.J., COUTAVAS E., BLOBEL G.;
 RT "A novel ubiquitin-like modification modulates the partitioning of
 RT the Ran-GTPase-activating protein RanGAP1 between the cytosol and the
 RT nuclear pore complex.";
 RL J. CELL BIOL. 135:1457-1470(1996).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC SPECIES-HUMAN; TISSUE-PLACENTA;

RX MEDLINE; 97064180.
 RA OKURA T., GONG L., KAMITANI T., WADA T., OKURA I., WEI C.F.,
 RA CHANG H.M., YEH E.T.H.;
 RT "Protection against Fas/Apo-1- and tumor necrosis factor-mediated
 RT cell death by a novel protein, sentrin.";
 RL J. IMMUNOL. 157:4277-4281(1996).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC SPECIES-MOUSE; STRAIN-ICR;
 RX MEDLINE; 98126440.
 RA HOWE K., WILLIAMSON J., BODDY M.N., SHEER D., FREEMONT P.S.,
 RA SOLOMON E.;
 RT "The ubiquitin-homology gene PIC1: characterization of mouse (Pic1)
 RT and human (UBL1) genes and pseudogenes.";
 RL GENOMICS 47:92-100(1998).
 RN [8]
 RP STRUCTURE BY NMR.
 RC SPECIES-HUMAN;
 RX MEDLINE; 98319859.
 RA BAYER P., ARNDT A., METZGER S., MAHAJAN R., MELCHIOR F., JAENICKE R.,
 RA BECKER J.;
 RT "Structure determination of the small ubiquitin-related modifier
 RT SUMO-1.";
 RL J. MOL. BIOL. 280:275-286(1998).
 CC -1- FUNCTION: ASSOCIATES WITH RAD51/RAD52. INVOLVED IN TARGETING
 CC RANGAP1 TO THE NUCLEAR PORE COMPLEX PROTEIN RANBP2.
 CC -1- SIMILARITY: TO UBIQUITIN. BELONGS TO THE SMT3 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X9586; E274603; -;
 DR EMBL; U61397; G1518694; -;
 DR EMBL; U38784; G1574848; -;
 DR EMBL; U67122; G1762873; -;
 DR EMBL; U72722; G1703503; -;
 DR EMBL; U83117; G1769602; -;
 DR EMBL; AF033353; G2645737; -;
 DR MIM; 601912; -;
 DR PDB; 1A5R; 1A-OCT-98.
 DR PROSITE; P550053; UBIQUITIN_2; 1.
 RW 3D-STRUCTURE.
 FT DOMAIN 20
 FT SEQUENCE 101 AA; 11557 MW; EC25CIFE CRC32;

Query Match 5.2%; Score 149; DB 1; Length 101;
 Best Local Similarity 28.6%; Pred. No. 4.07e-06;
 Matches 20; Conservative 23; Mismatches 26; Indels 1; Gaps 1;

DB 22 IRLKVIQDSSSEI-HFRVYKMTLKKLESYQCGRGVPMNSIRFLFEGORADNHTPREL 80
 QY 341 LRLRVQGEKHEHMLEISLSPDSPKLKMSHYEAMGLSGHKLSFFDGTUSGKELPADL 400
 DB 81 GMEEDVIEV 90
 QY 401 GLESGDLLEV 410

RESULT 6
 ID SM31_HUMAN STANDARD; PRT; 103 AA.
 AC P55854;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE UBIQUITIN-LIKE PROTEIN SMT3A.
 SMT3H1 OR SMT3A.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;

OC PRIMATES: CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-BRAIN.
MEDLINE: 97237059.
RA LAPERTA V., CHIBRAZI P., VAN DER SPEK P.J., PIZZUTI A.,
HANAKA F., BRAHE C.;
RT "SMTA, a human homologue of the S. cerevisiae SMT3 gene, maps to
chromosome 21qter and defines a novel gene family."
RL GENOMICS 40:362-367(1997).
CC -1- SIMILARITY: TO UBIQUITIN. BELONGS TO THE SMT3 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X9584; E274634; -.
CC DR MIM: 602231; -.
CC DR PROSITE: PSS0053; UBIQUITIN-2; 1.
CC FT DOMAIN 15 92 UBIQUITIN-LIKE.
CC SQ SEQUENCE 103 AA: 11654 MW: 6E303BC0 CRC32:
CC
Query Match 4.9%; Score 141; DB 1; Length 103;
Best Local Similarity 30.4%; Pred. No. 5,10e-05;
Matches 24; Conservative 20; Mismatches 33; Indels 2; Gaps 2;
DB 8 EGYPTENDHINKAYAGDGS-VYGFKIKRHTSLSKMKAYCERGLSMRQIRFEDQPI 66
QY 333 EAEETSOE-LRLRYGKEKHOMELISLSPSPKLVMSHYEANGLSGKLSFFEDGTL 391
DB 67 NETDTPAQLMEDDEDITDV 85
QY 392 SGRKLPADLGESGLIEV 410
CC
RESULT 7
ID IF4G_RABIT STANDARD; PRT; 1402 AA.
AC P41110;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE EUKARYOTIC TRANSLATION INITIATION FACTOR 4 GAMMA (EIF-4-GAMMA) (EIF-4G)
DE (EIF4G) (P220).
DE EIF4G.
OS OREOZYGUS CUNICULUS (RABBIT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC LAGOMORPHA; LEPORIDAE; ORCTOLAGUS.
RN [1]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 479-500.
RX STRAIN-NEW ZEALAND WHITE; TISSUE-BRAIN;
RX MEDLINE: 93374895.
RA LAMPHAR B.J., YAN R., YANG F., WATERS D., LIEBIG H.-D.,
KLUMP H., KOEHLER E., SKERN T., RHOADS R.E.;
RT "Mapping the cleavage site in protein synthesis initiation factor
EIF-4 gamma of the 2A proteases from human Cocksackievirus and
rhinovirus".
RL J. BIOL. CHEM. 268:19200-19203(1993).
CC [2]
CC PARTIAL SEQUENCE.
CC RP MEDLINE: 93054654.
CC RA YAN R., RYCHLIK W., ETCHEISON D., RHOADS R.E.;
CC RT "Amino acid sequence of the human protein synthesis initiation factor
EIF-4 gamma".
CC J. BIOL. CHEM. 267:23226-23231(1992).
CC -1- FUNCTION: COMPONENT OF THE PROTEIN COMPLEX EIF-4, WHICH IS
CC INVOLVED IN THE RECOGNITION OF THE MRNA CAP, ATP-DEPENDENT
CC UNWINDING OF 5'-TERMINAL SECONDARY STRUCTURE AND RECRUITMENT OF
CC MRNA TO THE RIBOSOME.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RNP).

CC -1- SIMILARITY: THE C-TERMINAL REGION IS SIMILAR TO THE N-TERMINAL
CC REGION OF WHEAT EUKARYOTIC INITIATION FACTOR (ISO)4F SUBUNIT P82.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L22090; G404775; -.
CC DR PROSITE: PSS00030; RNP-1; FALSE_NEG.
CC KW INITIATION FACTOR; PROTEIN BIOSYNTHESIS; PHOSPHORYLATION;
CC KW RNA-BINDING.
CC FT DOMAIN 188 192 POLY-PRO.
CC FT DOMAIN 262 275 POLY-GLU.
CC FT DOMAIN 602 607 RNA-BINDING (RNP2) (BY SIMILARITY).
CC FT DOMAIN 698 706 RNA-BINDING (RNP1) (BY SIMILARITY).
CC FT DOMAIN 1393 1398 POLY-GLU.
CC SQ SEQUENCE 1402 AA: 154050 MW: 6B00C346 CRC32:
CC
Query Match 4.5%; Score 129; DB 1; Length 1402;
Best Local Similarity 20.8%; Pred. No. 1,98e-03;
Matches 46; Conservative 62; Mismatches 103; Indels 10; Gaps 10;
DB 482 ALSSRGP-PRGPGGELPRGAGLGRRSLOPPRGARKLIVMTEDIKLNKAKAW 540
QY 4 PLRGGRPRSGRGARRARGRCPRARQSPARLPDVLVDVSDSEVLVADVPE 63
DB 541 KPSSKRRADK-DR-GEADGSKTODLFRVRSILNKLPQFQOLMKOYLAIDTEG 598
QY 64 VPAARLPAPAKPPQDDSDSEGA-AGCAPAPRTLVRRRRRLDGEAPVYV-YSGRV 121
DB 599 ASKGSILSSLRPFQNPPTSCQHPVLPHGASATTEKPTVTVN-FRKLNRCKEFE 657
QY 122 QSSLNLPDSSILKLPSPF-PEDEADLTNGSSPSBEDALPGSGWRKLRKCKE-E 179
DB 658 KDRDDEVEFEKKQEMDEATAERERLKELEARDIARR 698
QY 180 K-KME-EFPQDISPLPQPSRRKSRKHTALQKLEVNRR 218
CC
RESULT 8
ID SMT3_SCHPO STANDARD; PRT; 90 AA.
AC O13351;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE UBIQUITIN-LIKE PROTEIN SMT3.
DE SMT3.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-358;
RX PELLETIER M.F., DIGNARD D.;
RT SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RL -1- SIMILARITY: TO UBIQUITIN. BELONGS TO THE SMT3 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF019235; G2444268; -.
CC DR PROSITE: PSS0053; UBIQUITIN-2; 1.
CC FT DOMAIN 7 87 UBIQUITIN-LIKE.

NAT. GENET. 3:137-145(1993).

CC -1- FUNCTION: UNKNOWN.

CC -1- TISSUE SPECIFICITY: LIMITED TO CELL-LINES OF LEUKEMIC ORIGIN.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M33509; G179339; -

DR EMBL; M33518; G179345; -

DR EMBL; M33512; G179345; JOINED.

DR PIR; B35098; B35098.

DR PIR; S36152; S36152.

DR MIM; 142580; -

MM REPEAT.

FT DOMAIN 519 524 POLY-PRO.

FT DOMAIN 636 657 GLN-RICH.

FT DOMAIN 684 688 POLY-PRO.

FT DOMAIN 699 704 POLY-PRO.

FT DOMAIN 814 821 POLY-PRO.

FT DOMAIN 1340 1345 POLY-GLY.

FT DOMAIN 1398 1403 POLY-GLY.

FT DOMAIN 1436 1442 POLY-PRO.

FT DOMAIN 1982 1991 POLY-PRO.

FT REPEAT 41 95 4 X 57 AA TYPE A REPEATS.

FT REPEAT 98 154 1-2.

FT REPEAT 281 337 1-3.

FT REPEAT 1740 1795 1-4.

FT DOMAIN 337 549 2 X TYPE B REPEATS.

FT REPEAT 337 418 2-1.

FT REPEAT 476 549 2-2.

FT DOMAIN 1899 2089 3 X 50 AA TYPE C REPEATS.

FT REPEAT 1899 1948 3-1.

FT REPEAT 1965 2014 3-2.

FT REPEAT 2040 2089 3-3.

FT REPEAT 57 57 R -> A (IN REF. 2).

FT REPEAT 109 109 Q -> S (IN REF. 2).

FT REPEAT 414 414 P -> P (IN REF. 2).

FT REPEAT 532 532 T -> K (IN REF. 2).

FT REPEAT 682 682 Q -> K (IN REF. 2).

FT REPEAT 730 730 E -> D (IN REF. 2).

FT REPEAT 750 750 L -> R (IN REF. 2).

FT REPEAT 834 834 A -> T (IN REF. 2).

FT REPEAT 1035 1035 G -> A (IN REF. 2).

FT REPEAT 1068 1068 M -> L (IN REF. 2).

FT REPEAT 1285 1285 P -> R (IN REF. 2).

FT REPEAT 1400 1400 G -> A (IN REF. 2).

FT REPEAT 1611 1611 T -> S (IN REF. 2).

FT REPEAT 1729 1729 G -> A (IN REF. 2).

SEQUENCE 2142 AA; 227840 MW; 2CFEF88A CRC32;

Query Match 4.0%; Score 113; DB 1; Length 2142;

Best Local Similarity 29.5%; Pred. No. 1.93e-01;

Matches 31; Conservative 30; Mismatches 38; Indels 6; Gaps 6;

DB 1012 RGRGEYFARG-RGRGTYGRRGGGANSATVESEFEMMGVE-VGGGDTTLLLEPA-M 1068

QY 6 RGRGP-RSRGGGARRARRARRCRPARLRIPDTLVLDVSDSEVLEVADPVEV 64

DB 1069 P-ARHGARVQSMKSPSGAGGAKOAKAPMVRVIMLQTRRLPHR 1112

QY 65 PVARLPAPAKPEODSDSDEGAEGPAGPRLVRR-RRRRLDP 108

RESULT 12 STANDARD; PRT; 359 AA.

ID ISL2 BRARE

AC P53406;

DT 01-OCT-1996 (REL. 34, CREATED)

DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)

DE INSULIN GENE ENHANCER PROTEIN ISL-2 (ISLET-2).

GN ISL2 OR ISL-2.

OS BRACHDANIO BERTIO (ZEBRAFISH) (ZEBRA DANIO).

CC EUDAROTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;

CC TELEOSTEI; EUTELEOSTEI; OSTARIOPTYSII; CYPRINIFORMES; CYPRINOIDEA;

CC CYPRINIDE; RASBORINAE; DANIO.

CC [1]

RN SEQUENCE FROM N.A.

RC TISSUE-EMBRYO;

RA MEDLINE; 96005022.

RA TOKUMOTO M., GONG Z., TSUBOKAWA T., HEW C.L., UYEMURA K., HOTTA Y.,

RA OKAMOTO H.;

RT "Molecular heterogeneity among primary motoneurons and within

RT myotomes revealed by the differential mRNA expression of novel

RT islet-1 homologs in embryonic zebrafish."

RL DEV. BIOL. 171:578-589(1995).

RN [2]

RP SEQUENCE FROM N.A.

RA MEDLINE; 96125154.

RA APPEL B., KORZ V., GLASGOW E., THOR S., EDLUND T., DAVID L.,

RA EISEN J.;

RT "Motoneuron fate specification revealed by patterned LIM homeobox

RT gene expression in embryonic zebrafish."

RL DEVELOPMENT 121:4117-4125(1995).

CC -1- FUNCTION: BINDS TO ONE OF THE CIS-ACTING DOMAIN OF THE INSULIN

CC GENE ENHANCER. MAY BE INVOLVED IN SUBTYPE SPECIALIZATION OF

CC PRIMARY MOTONEURONS.

CC -1- SUBCELLULAR LOCATION: NUCLEAR.

CC -1- DEVELOPMENTAL STAGE: FIRST EXPRESSED AFTER 15 H OF GASTRULATION,

CC SEGMENTALLY IN THE VENTRAL REGION OF THE SPINAL CORD AND THEN

CC LATER IN THE DORSAL REGIONS (ROHAN-BEARD NEURONS).

CC -1- SIMILARITY: TO OTHER HOMEOBOX DOMAINS. BELONGS TO THE LIM

CC SUBFAMILY.

CC -1- SIMILARITY: CONTAINS 2 LIM DOMAINS. THE LIM DOMAIN BINDS 2 ZINC

CC IONS.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; D38453; G1037166; -

DR EMBL; U09403; G487805; -

DR EMBL; X88803; G4871001; -

DR PROSITE; PS00027; HOMEOBOX_1; 1.

DR PROSITE; PS00478; LIM_DOMAIN_1; 2.

DR PROSITE; PS50023; LIM_DOMAIN_2; 2.

DR PROSITE; PS50071; HOMEOBOX_2; 1.

DR PFM; PF00045; homeobox; 1.

DR PFM; PF00412; LIM; 2.

DR HSSP; P02836; 1HDD.

DR HOMEOBOX; DNA-BINDING; DEVELOPMENTAL PROTEIN; NUCLEAR PROTEIN;

DR REPEAT; LIM MOTIF; METAL-BINDING; ZINC; MULTIGENE FAMILY.

FT DOMAIN 27 80 LIM.

FT DNA_BIND 30 143 LIM.

FT DOMAIN 191 250 HOMEOBOX.

FT DOMAIN 255 359 GLN-RICH.

SEQUENCE 359 AA; 40227 MW; DC592CD0 CRC32;

Query Match 3.9%; Score 111; DB 1; Length 359;

Best Local Similarity 30.3%; Pred. No. 3.33e-01;

Matches 23; Conservative 23; Mismatches 24; Indels 6; Gaps 6;

DB 174 PVROPFRHHVHKOSEKTRVTLNKKOHLTLRTCYVANRPD-ALM-KE-QLEWMTG- 229

QY 192 PLPOSSRNKRKHTLAKLREV-N-KRLDLSCLSPKQHSBALDSTDEYVLVSGP 249

```
D6      230    LSPRVIRWPFONKRCR   245
        :| |: |:: ||:
OY       250    VLPQSSRLFTLKIKCR   265


RESULT_13
ID VE2_HPV37          STANDARD:             PRT;         454 AA.
AC Q80903:
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DE 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DR REGULATORY PROTEIN E2.
GN E2.
OS HUMAN PAPILLOMAVIRUS TYPE 37.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; PAPOVAVIRIDAE; PAPILLOMAVIRUS.
RN (1)
RP SEQUENCE FROM N.A.
RA DELIUS H.;
RL SUBMITTED (OCT-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC -! FUNCTION: E2 REGULATE THE ESSENTIAL TRANSCRIPTION AND DNA REPLICATION
IT BINDS TO THE E2E RESPONSE ELEMENT (5'-ACCNNNNNGGT-3') PRESENT
IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF EBV'S POSITION
WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSOR OCCURS
BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
INITIATION COMPLEX. THE EI-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
REPLICATION.
CC -! SUBCELLULAR LOCATION: NUCLEAR.
CC -! SUBUNIT: BINDS DNA AS A DIMER.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation at
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isdb-sib.ch/announce/
or send an email to license@sib-sib.ch).
-----
D6      EMBL; U31786; G1020230; -.
DR PFAM; PF00508; E2_N; 1.
DR PFAM; PF00511; E2_C; 1.
DR HSPP; P17383; IDHM.
KW EARLY PROTEIN; TRANSCRIPTION REGULATION; ACTIVATOR; DNA-BINDING;
TRANS-ACTING FACTOR; DNA REPLICATION; REPRESSOR; NUCLEIC ACID-BINDING
SEQUENCE     454 AA; 51285 MW; EA93070B CRC32;
SQ
Query March              3.9%; Score 112; DB 1; Length 454;
Best local similarity    44.7%; Pred No 2.53e+01;
Matches 21; Conservative 13; Mismatches *            8; Indels 5; Gaps 4;

300 SSPDRGGRSGRGPEPSTRQSRSLSRSRSRSRSSGSSTGGAGPADAV 346
:||||||| |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
2 AEPHGRRGPBRSGRGAR-RARG-ARGC-PARRASPAR-LIPPTV 43


RESULT_14
ID RHO_STRLI          STANDARD:             PRT;         707 AA.
AC P52157:
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE TRANSCRIPTION TERMINATION FACTOR RHO.
GN RHO.
OS STREPTOMYCETES LIVIVANS.
OC BACTERIA: FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIIDAE;
ACTINOMYCETALES; STREPTOMICINEAE; STREPTOMICETACEAE; STREPTOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-Zx7;
RA INGHAM C.J., HUNTER I.S., SMITH M.C.M.;
BL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CL -! FUNCTION: FACILITATES TRANSCRIPTION TERMINATION BY A MECHANISM
THAT INVOLVES RHO BINDING TO THE NASCENT RNA, ACTIVATION OF RHO'S
```

```

CC RNA-DEPENDENT ATPASE ACTIVITY, AND RELEASE OF THE MRNA FROM THE
CC DNA TEMPLATE (BY SIMILARITY).
CC -1- SUBUNIT: HOMOHXAMER (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RNP).
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-43 IS THE INITIATOR.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
CC DR EMBL; X95444; E220353; -.
CC DR EMBL; X95444; E220354; ALT_INIT.
CC DR HSSP; P03002; 1A62.
CC KW TRANSCRIPTION TERMINATION; HELICASE; ATP-BINDING; RNA-BINDING.
CC FT DOMAIN 38 43 RNA-BINDING (RNP2) (BY SIMILARITY).
CC FT DOMAIN 343 346 RNA-BINDING (RNP1) (BY SIMILARITY).
CC FT NP_BIND 458 465 ATP (POTENTIAL).
CC FT DOMAIN 33 36 POLY-ARG.
CC FT DOMAIN 168 172 POLY-ARG.
CC FT DOMAIN 280 283 POLY-GLY.
CC SQ SEQUENCE 707 AA; 76544 MW; BC013166 CRC32;
CC
CC Query Match 3.9%; Score 112; DB 1; Length 707;
CC Best Local Similarity 48.5%; Pred. No. 2,53e+01;
CC Matches 16; Conservative 3; Mismatches 13; Indels 1; Gaps 1;
CC
CC Db 230 RDRGDRDGRDGRDGRDGRDGRDGRDGRGDRRKK 262
CC | | | | | | | | | | | | | | | | | | |
CC Qy 6 RGRGRSRGGRGARRARG-RGPRARQSPAR 37
CC
CC RESULT 15
CC ID FBPI_DROME STANDARD; PRT; 1030 AA.
CC AC Q04691;
CC DT 01-JUN-1994 (REL. 29, CREATED)
CC DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
CC DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
CC DE FAT-BODY PROTEIN-1 PRECURSOR (P1 PROTEIN).
CC GN FBPI OR P1.
CC OS DROSOPHILA MELANOGASTER (FRUIT FLY).
CC OC EURARCTOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
CC CC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYROIDEA;
CC CC DROSOPHILIDAE; DROSOPHILA.
CC [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN-CANTON-S; TISSUE=FAT BODY;
CC RC MEDLINE; 90339481.
CC RA MASCHAF F., DUBERTRET M.-L., THEONOD P., CLAVERIE J.-M.,
CC RA LEPESSANT J.-A.,
CC RT "Structure of the ecdysone-inducible P1 gene of Drosophila
CC RT melanogaster.";
CC RT J. MOL. BIOL. 214:359-372(1990).
CC [2]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN-CANTON-S;
CC RC MEDLINE; 94123973.
CC RA LARIE P., NASR F., LEPESSANT J.-A., DEUTSCH J.;
CC RT "Deletion scanning of the regulatory sequences of the Fbpi1 gene of
CC RT Drosophila melanogaster using P transposase-induced deficiencies.";
CC RL GENETICS 135:801-816(1993).
CC CC -1- TISSUE SPECIFICITY: FAT BODY.
CC CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING THE LATE THIRD LARVAL
CC CC STAGE.
CC CC -1- INDUCTION: BY ECDYSONE.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X54997; G8320; -
DR EMBL: X69965; G7961; -
DR PIR: S1034; S1034; -
DR FLYBASE: FBgn000639; Fbpl.
DR PFAM: PF00372; hemocyanin; 2.
KM SIGNAL: GLYCOPROTEIN.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 1030 FAT-BODY PROTEIN-1.
FT DOMAIN 360 371 POLY-ASP.
FT DOMAIN 580 590 POLY-ASP.
FT CARBOHYD 741 741 POTENTIAL.
SO SEQUENCE 1030 AA; 119365 MW; A3AC5F31 CRC32;

Query Match 3.9% Score 111; DB 1; Length 1030;
Best Local Similarity 20.9% Pred. No. 3.33e-01;
Matches 42; Conservative 66; Mismatches 76; Indels 17; Gaps 15;

Db 559 TDDEILEMLORRROQLQKH-QNDDDD-DD-NDDVNVVHROGLSRSRSLPRLRQNNRL 615
QY 156 SEDDALPS-GSPWRKKLKKCKEKKMEFPDIDISPLPQSSRNKSRKHTALQKLE 214
Db 616 SEIVLHNRQ-LVARLNQESIAQGLIEQQQLINPRLTOSERYALRLNOIRINSQSR 674
QY 215 VNKRLQDLRSCLSPKQHSPALQS--TDDEVVLVEGPLYPOSRL-FTLK-IRCRADLYR 270
Db 675 QVLAQIGIEQRIQVIGOVLSOVVNSLROGVLDQROVESLIADVLGRLGQVIGIMTII 734
QY 271 --LP-V-RMSEPLQVVDHMANHLGVSPNRIILFG-ESE-LSPATPSTL-KLSVADII 323
Db 735 RQVVDNNSQIDRNGLGIRL 755
QY 324 DCVVLASSSEATE-TSQELRL 343

Search completed: Fri Feb 18 11:59:27 2000
Job time : 51 secs.

Db 301 FGESELSPTATPSTLKLGVADIIIDCVIASSSEATETSOELRLRVQKQKHOMLEISLSP 360
 QY 301 FGESELSPTATPSTLKLGVADIIIDCVIASSSEATETSOELRLRVQKQKHOMLEISLSP 360
 Db 361 DSPILKVMSHYEFAMGLSGHKLSPFFDGTKLSGKELPADLGLSGDLIEVWG 412
 QY 361 DSPILKVMSHYEFAMGLSGHKLSPFFDGTKLSGKELPADLGLSGDLIEVWG 412

RESULT 2
 ID 057686 PRELIMINARY; PRT: 102 AA.

AC 057686; 01-JUN-1998 (TREMBLREL. 06, CREATED)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE SUMO-1 PROTEIN.
 GN SUMO-1.
 OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
 OC EUKARYOTA; METAFOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
 CC MESOBATRYCHIA; PIPOIDEA; PIPOIDAE; XENODIDINAE; XENOPUS.
 RN [1]
 RX MEDLINE: 98089176.
 RA SATOH H., SPARROW D.B., SHIOMI T., PU R.T., NISHIMOTO T., MOHUN T.J.,
 RA DASSO M.,
 RT "Dc9p and the conjugation of SUMO-1 to Rangap1 and Ranbp2.";
 RL CURR. BIOL. 8:121-124(1998).
 DR EMBL: 297073; E323533; -
 SQ SEQUENCE 102 AA; 11703 MW; F3622FE3 CRC32;

Query Match 5.6%; Score 161; DB 13; Length 102;
 Best Local Similarity 26.6%; Pred. No. 4.33e-07;
 Matches 21; Conservative 27; Mismatches 30; Indels 1; Gaps 1;

Db 14 GDKKDGDIYIKLVIGODSSEI-HFKVMTLTKLKESYRQGVPMNSLRFEGORI 72
 QY 332 SEATETSOELRLRVQKQKHOMLEISLSPDSPILKVMSHYEFAMGLSGHKLSPFFDGTKL 391
 QY 392 SGKELPADLGLSGDLIEV 410

RESULT 3
 ID 023759 PRELIMINARY; PRT: 115 AA.

AC 023759; 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 DE UBIQUITIN PROTEIN.
 OS CICCER ARIETINUM (CHICKPEA) (GARBANZO).
 OC EUKARYOTA; VIRIDIPHYTES; STREPTOPHYTA; EMERYOPHYTA; TRACHEOPHYTA;
 CC EUPHYLYPHITES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
 OC FABALES; FABACEAE; PAPILIONOIDEAE; CICCER.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CY. CASTELLANA; TISSUE=ETIOLATED EPICOTYLS;
 RA MUNOZ F.J., DOPICO B., LABRADOR E.;
 RL PLANT PHYSIOL. 116:1605-1605(1998).
 DR EMBL: AJ001901; E354254; -
 SQ SEQUENCE 115 AA; 12858 MW; 89C22D7C CRC32;

Query Match 5.6%; Score 161; DB 10; Length 115;
 Best Local Similarity 25.0%; Pred. No. 4.33e-07;
 Matches 19; Conservative 30; Mismatches 25; Indels 2; Gaps 2;

Db 18 PNDAAH-INLKVGGDGNEVF-FRIKRTQKLKLNATCDQSDVLSIAFLFGRRIRA 75
 QY 334 ATETSOELRLRVQKQKHOMLEISLSPDSPILKVMSHYEFAMGLSGHKLSPFFDGTKL 393
 Db 76 EOTPDLEMDGDEID 91
 QY 394 KELPADLGLSGDLIE 409

RESULT 4
 ID 057114 PRELIMINARY; PRT: 239 AA.

AC 057114; 01-JUN-1998 (TREMBLREL. 06, CREATED)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 DE NONSTRUCTURAL PROTEIN P125-2 (FRAGMENT).
 GN P125.
 OS PESTIVIRUS TYPE 1.
 CC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES. NO DNA STAGE; FLAVIVIRIDAE;
 CC PESTIVIRUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SNC;
 RA OT F., BERRY E.S.;
 RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: U89438; G2707597; -
 KW NONSTRUCTURAL PROTEIN.
 FT NON TER 1
 FT NON TER 1
 SQ SEQUENCE 239 AA; 26881 MW; 984208C6 CRC32;

Query Match 5.6%; Score 161; DB 14; Length 239;
 Best Local Similarity 32.9%; Pred. No. 4.33e-07;
 Matches 25; Conservative 19; Mismatches 31; Indels 1; Gaps 1;

Db 68 TENNDHNLNVAAGDGS-VVOFKIKRTPLSKLMKACERQGSMDRIIRFQSGPINET 126
 QY 335 TETSOELRLRVQKQKHOMLEISLSPDSPILKVMSHYEFAMGLSGHKLSPFFDGTKL 394
 Db 127 DTPAQLEMEDEDITDV 142
 QY 395 ELFPADLGLSGDLIEV 410

RESULT 5
 ID 074186 PRELIMINARY; PRT: 117 AA.

AC 074186; 01-NOV-1998 (TREMBLREL. 08, CREATED)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE PMT3P.
 GN PMT3.
 OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
 CC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
 CC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA NISHIDE J., NISHI T., MURAKAMI Y., MATSUDA H., KAWAMUKAI M.,
 RA TANAKA K.;
 RT "The fission yeast ubiquitin-like modifier.";
 RL SUBMITTED (AUG-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AB017187; D1033560; -
 SQ SEQUENCE 117 AA; 12934 MW; 9807EC2C CRC32;

Query Match 4.6%; Score 131; DB 3; Length 117;
 Best Local Similarity 25.9%; Pred. No. 3.41e-03;
 Matches 21; Conservative 24; Mismatches 35; Indels 1; Gaps 1;

Db 24 TSOQDVNPSTEHLNKLKVVAGDNEVF-FRIKRTTFSKLKIKCARQKSMNSLRLVLDG 82
 QY 329 ASSEATETSOELRLRVQKQKHOMLEISLSPDSPILKVMSHYEFAMGLSGHKLSPFFDGT 388
 Db 83 ERIRPDQTPALDMDGQDTE 103
 QY 389 TKLSGKELPADLGLSGDLIE 409

RESULT 6
 ID 063624 PRELIMINARY; PRT: 1173 AA.

[illegible][illegible]

THIS PAGE BLANK (USPTO)